



Monitoring mean and variability by Gini chart for skew-normal distributed data

M. Sasaei, R. Pourmoussa, and M. Mashinchi*

Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.

Received 15 October 2019; received in revised form 30 April 2020; accepted 8 June 2020

KEYWORDS

Control charts;
 False alarm rate;
 Gini chart;
 Skew-normal
 distribution;
 Power curves.

Abstract. In order to monitor the mean and variability of a process, the Gini control charts are proposed based on the skew normally distributed random sample. By comparing the false alarm rates of current scheme with those of existing mean and dispersion control charts, it was found that the design structure of Gini chart could be enhanced in comparison to other classic schemes based on the assumption of skew normal distribution for the data. Moreover, the superiority of the Gini chart was investigated by comparing the discriminatory power curves of the skew normal distribution with some existing control charts. Simulated studies and a real data example illustrate the usefulness of the proposed approach.

© 2022 Sharif University of Technology. All rights reserved.

1. Introduction

Statistical quality control techniques such as control charts, which are commonly used in the industrial sector, are used to determine the control limits and explore the changes that are required to make for the purpose of improving the process. The Conventional Shewhart (CS) control chart is a key tool for process improvement in order to detect the occurrence of assignable causes such that quality practitioners can take necessary corrective actions in the early stage of manufacturing when a large quantity of non-conformities are produced [1]. Although the Shewhart-type control chart is based on the assumption of data normality, it is widely believed that many manufacturing processes do not fulfill the normality assumption and its property. The fat tails as well as peaked and skewed features are often the most important non-normality properties in the industrial

data. For instance, the studies [2–5] suggested that the cutting tool wear, chemical, and semiconductor processes followed a skewed distribution. Therefore, the application of the CS control chart without considering the effect of skewness in data increases the false alarm rate and may yield misleading results. The growth in the false alarm rate in the presence of skewness is rooted in the discrepancy between the variability pattern of the asymmetric distribution and the normality assumption in the construction of control charts [5]. Therefore, in [5], a Weighted Standard Deviation (WSD) method was proposed and it functioned based on the idea that a skewed distribution could be bisect at its mean and two new symmetric distributions with the same mean. However, different standard deviations were formed by each segment of the original skewed distribution so that they could be applied for setting up the limits of the control chart. Later, this research study [6] proposed a Skewness Correction (SC) method based on the idea of using Cornish-Fisher expansion with emphasis on the degree of skewness of the process distribution.

Gini's mean difference, initially introduced by Gini in 1912, has received much attention in measuring the variability of data in a large number of scientific

*. Corresponding author.

E-mail addresses: mahsasasaei@gmail.com (M. Sasaei);
pourm@uk.ac.ir (R. Pourmoussa); mashinchi@uk.ac.ir (M. Mashinchi)

applications and research area. It is broadly used and considered as a suitable base for developing control charts. For instance, see the recent works of [7–20] among the others and the acknowledged references therein. By using Gini's mean difference estimate as a measure of dispersion, Saghir and Lin [14] derived the Gini charts for the non-normal models such as exponential, student-t, Logistic, and Laplace distributions. Moreover, the well-known Skew-Normal (SN) distribution [21] was used for statistically developing control charts. In [22], the \bar{X} control chart of SN process was constructed by utilizing the approximate distribution of the sample mean. Later, in [1], \bar{X} and R control charts were developed for skew normally distributed data by obtaining their exact distributions. This paper [1] compared the performance of their methodology by the shewhart control charts with the structures in [5,6,22]. It was demonstrated that their new control charts could offer considerable improvement over existing methods.

The main objective of this paper is to develop SN control charts based on the Gini's mean difference measure and to derive the efficiency of using the Gini estimate relative to the sample range R as a measure of dispersion in design structure of control charts. Therefore, the rest of this paper is organized as follows. Section 2 summarizes some preliminaries on Li's control charts methodology by SN distribution and Gini control charts in general. In Section 3, the Gini control charts for the SN distribution are developed to monitor the process mean and variability. In order to demonstrate the effectiveness of the proposed control charts, some numerical examples, including both simulated and real data, are presented in Sections 4 and 5. Finally, Section 6 provides some concluding remarks.

2. Preliminaries

2.1. The SN distribution and its control charts

Suppose that the quality characteristic X follows the SN distribution with the location, scale, and skewness parameters ξ , σ , and λ , illustrated by $X \sim SN(\xi, \sigma, \lambda)$, respectively.

The probability density function (pdf) of X is:

$$f_X(x; \xi, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x - \xi}{\sigma}\right) \Phi\left(\lambda \frac{x - \xi}{\sigma}\right),$$

$$\xi, \lambda \in \mathbb{R}, \quad \sigma \in \mathbb{R}^+, \quad (1)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and the cumulative density function (cdf) of the standard normal distribution, respectively [21]. It can be seen that λ controls the shape of the distribution and the SN distribution coincides with the normal one with mean ξ and variance σ^2 , as λ tends to zero. If $\lambda > 0$, the SN distribution is

positively skewed; in the case of $\lambda < 0$, it is negatively skewed. It is also quite clear that the mean and standard deviation of the SN distribution are $\mu_x = \xi + \sigma\gamma_2/\gamma_1$ and $\sigma_x = \sigma/\gamma_1$ where $\gamma_1 = (1 - a_1^2\rho^2)^{-1/2}$, $\gamma_2 = \gamma_1 a_1 \rho$, $a_1 = \sqrt{2/\pi}$, and $\rho = \lambda/\sqrt{1 + \lambda^2}$. We will denote the standard SN distribution, $\xi = 0$ and $\sigma = 1$, hereafter by $SN(\lambda)$.

Owing to the proven proficiency and properties of the SN distribution against normal distribution [23,24], two straightforward approaches of control limits, proposed by Li et al. [1] and Tsai [22], are available. Recently, by using the exact distributions of \bar{X} and R as the sample mean and sample range, respectively, Li et al. [1] obtained the SN \bar{X} and R control limits as:

SN \bar{X} chart :

$$\begin{cases} LCL_{\bar{X}} = \mu_{x_0} - (\gamma_2 - \gamma_1 \bar{z}_{\lambda, n, \alpha/2}) \sigma_{x_0} / \sqrt{n}, \\ UCL_{\bar{X}} = \mu_{x_0} + (\gamma_1 \bar{z}_{\lambda, n, 1-\alpha/2} - \gamma_2) \sigma_{x_0} / \sqrt{n}, \end{cases}$$

SN R chart :

$$\begin{cases} LCL_R = \gamma_1 r_{\lambda, n, \alpha/2} \sigma_{x_0}, \\ UCL_R = \gamma_1 r_{\lambda, n, 1-\alpha/2} \sigma_{x_0}, \end{cases} \quad (2)$$

where μ_{x_0} and σ_{x_0} are the in-control process mean and standard deviation, respectively, which should be estimated if they are unknown. Assuming that a dataset is drawn from $SN(\lambda)$, $\bar{z}_{\lambda, n, \alpha}$ in Eq. (2) denotes the quantile of the sample mean distribution \bar{Z} , i.e., $Pr(\bar{Z} < \bar{z}_{\lambda, n, \alpha}) = \alpha$, and $r_{\lambda, n, \alpha}$ the quantile of distribution of R .

To estimate μ_{x_0} and σ_{x_0} , let X_{ij} be the j th observation from the i th group, $i = 1, \dots, m$, $j = 1, \dots, n$, drawn from a SN distribution. Then, the estimate of process mean can be computed as the grand mean $\bar{\bar{X}} = \sum_{i=1}^m \sum_{j=1}^n X_{ij} / (nm)$. Moreover, σ_{x_0} can be estimated by $\bar{R}/(\gamma_1 d_2)$ where in the range of the i th group R_i , i.e., the range of X_{i1}, \dots, X_{in} , $\bar{R} = \sum_{i=1}^m R_i / m$ and d_2 is numerically computed through the exact distribution of the sample range [1].

Therefore, the control limits Eq. (2) change into the following:

SN \bar{X} chart :

$$\begin{cases} LCL_{\bar{X}} = \bar{\bar{X}} - (\gamma_2 - \gamma_1 \bar{Z}_{\lambda, n, \alpha/2}) \bar{R} / (\sqrt{n} \gamma_1 d_2), \\ CL_{\bar{X}} = \bar{\bar{X}}, \\ UCL_{\bar{X}} = \bar{\bar{X}} + (\gamma_1 \bar{Z}_{\lambda, n, 1-\alpha/2} - \gamma_2) \bar{R} / (\sqrt{n} \gamma_1 d_2), \end{cases}$$

SN R chart :

$$\begin{cases} LCL_R = \gamma_1 r_{\lambda, n, \alpha/2} \bar{R} / (\gamma_1 d_2), \\ CL_R = \bar{R}, \\ UCL_R = \gamma_1 r_{\lambda, n, 1-\alpha/2} \bar{R} / (\gamma_1 d_2). \end{cases} \quad (3)$$

2.2. Gini control charts

In order to monitor variability of a process, Shewhart gave two different control charts, namely R and S charts, back in the 1920s. The R chart loses its usefulness with an increase in sample size n and the S chart behaves badly in the presence of outliers in data [25]. To address this deficiency of S and R charts, following the pioneering work of [22] based on S chart and [1] based on R chart, we propose Gini chart based on Gini estimate of dispersion measure to monitor the changes in process variability when the data are generated from the SN distribution. In this regard, let X be the process quality characteristic and a random sample of size n as X_1, X_2, \dots, X_n is drawn from the population. Then, the Gini's sample mean difference is:

$$G = \sum_{i=1}^n \sum_{j=i+1}^n |X_i - X_j| / (n(n-1)/2). \quad (4)$$

Theoretically, finding the exact distribution of G is difficult and thus, a numerical technique such as Monte Carlo simulation is usually used. According to the study [7], another type of control limits is in the form of probability control limits, representing the equivalent substitute of three-sigma limits for symmetric processes and providing an appropriate fit under non-normal processes. Three-sigma control limits are suitable for a normal or at least symmetric process, but it eventually loses its effectiveness as the process deviates from symmetry, as stated by Abbasi and Miller [9]. In this paper, the probability control limits method is considered as an alternative way to point out the skewness-based structure of dispersion control charts for the monitoring process variability. Also, dispersion chart is seldom used singly and it should be combined with the chart for monitoring the location. The most commonly used control chart for monitoring the location parameter is \bar{X} chart. The next section is developed by the control limits of \bar{X} and dispersion chart where Gini's mean difference estimate is applied as dispersion measure for the SN distribution is considered.

3. The proposed method based on Gini chart for the SN process

3.1. \bar{X} control limits based on Gini's mean difference

Assume that X is the quality characteristic of the interest and let X_1, X_2, \dots, X_n be a set of random samples of size n that are identically distributed by $SN(\xi, \sigma, \lambda)$. Moreover, suppose the set Z_1, Z_2, \dots, Z_n is also taken from $SN(\lambda)$ distribution. For the predefined false alarm rate α , the control limits of \bar{X} chart, can be obtained as follows:

$$Pr(\bar{X} < LCL_{\bar{X}} | \mu_x = \mu_{x_0}, \sigma_x^2 = \sigma_{x_0}^2) = \frac{\alpha}{2},$$

$$Pr(\bar{X} > UCL_{\bar{X}} | \mu_x = \mu_{x_0}, \sigma_x^2 = \sigma_{x_0}^2) = \frac{\alpha}{2},$$

where μ_{x_0} and $\sigma_{x_0}^2$ are the in-control process mean and variance, respectively. Thereupon, considering the false alarm rate in (2.2) equals 0.0027, and following the Shewhart control chart limits, the control limits in (2.2) can be rewritten as follows:

$$\begin{aligned} LCL_{\bar{X}} &= \mu_{x_0} - A_1^L \sigma_{x_0} / \sqrt{n}, \\ UCL_{\bar{X}} &= \mu_{x_0} + A_1^U \sigma_{x_0} / \sqrt{n}, \end{aligned} \quad (5)$$

where $A_1^L = \gamma_2 - \gamma_1 \bar{z}_{\lambda, n, 0.00135}$, and $A_1^U = \gamma_1 \bar{z}_{\lambda, n, 0.99865} - \gamma_2$ [22]. Note that, $\bar{z}_{\lambda, n, \alpha}$ can be computed numerically through the distribution of \bar{Z} in [1].

Assuming the parameters are unknown, μ_{x_0} is usually estimated by the grand mean of the subgroup means $\bar{\bar{X}}$. If λ is unknown, it can be estimated by $[(b_1/\hat{k}_3)^{2/3} + a_1^2 - 1]^{-1/2}$ for $\hat{k}_3 \geq 0$, and by $[(-b_1/\hat{k}_3)^{2/3} + a_1^2 - 1]^{-1/2}$ for $\hat{k}_3 < 0$, where:

$$b_1 = (\pi/4 - 1)a_1,$$

$$\hat{k}_3 = \frac{1}{nm - 3} \sum_{i=1}^m \sum_{j=1}^n \left(\frac{X_{ij} - \bar{\bar{X}}}{S} \right)^3,$$

$$S = \sqrt{\frac{1}{nm - 1} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{\bar{X}})^2}.$$

Li et al. [1] defined $d_2 = E(R/(\sigma_{x_0} \gamma_1))$ and then derived the estimate of the process parameter σ_{x_0} as $\bar{R}/(\gamma_1 d_2)$. Similarly, we utilize sample Gini's mean difference in (4) to estimate σ_{x_0} in this article as $d_2^* = E(G/(\sigma_{x_0} \gamma_1))$. Then, the process parameter σ_{x_0} can be estimated by $\bar{G}/(\gamma_1 d_2^*)$, where, in which G_i is the Gini estimator of sample i , i.e., X_{i1}, \dots, X_{in} , in Eq. (4), and d_2^* can be computed by a Monte Carlo simulation procedure since G_i does not have an exact distribution when the quality variable is distributed by SN model. Therefore, the control limits of \bar{X} chart are obtained as in Eq. (6) if the parameters of the process are unknown:

SN Gini \bar{X} chart :

$$\begin{cases} LCL_{\bar{X}} = \bar{\bar{X}} - A_1^L \bar{G}/(\sqrt{n} \gamma_1 d_2^*) = \bar{\bar{X}} - A_2^L \bar{G}, \\ CL = \bar{\bar{X}}, \\ UCL_{\bar{X}} = \bar{\bar{X}} + A_1^U \bar{G}/(\sqrt{n} \gamma_1 d_2^*) = \bar{\bar{X}} + A_2^U \bar{G}, \end{cases} \quad (6)$$

where the charting constants are as $A_2^L = A_1^L/(\sqrt{n} \gamma_1 d_2^*)$ and $A_2^U = A_1^U/(\sqrt{n} \gamma_1 d_2^*)$.

3.2. Probability control limits based on Gini's mean difference

Let the sample Gini's mean difference of data drawn

from $SN(\xi, \sigma, \lambda)$ and $SN(\lambda)$ be G and G_Z , respectively. The probability control limits of the Gini chart under the given α are obtained as:

$$\begin{aligned} Pr(G < LCL_G | \sigma_x = \sigma_{x_0}) &= \frac{\alpha}{2}, \\ Pr(G > UCL_G | \sigma_x = \sigma_{x_0}) &= \frac{\alpha}{2}. \end{aligned} \quad (7)$$

Consequently, using the relationship $G = \gamma_1 \sigma_{x_0} G_Z$ based on definitions of G and G_Z , we can obtain:

$$\begin{aligned} Pr\left(G_Z < \frac{LCL_G}{\gamma_1 \sigma_{x_0}}\right) &= \frac{\alpha}{2}, \\ Pr\left(G_Z < \frac{UCL_G}{\gamma_1 \sigma_{x_0}}\right) &= 1 - \frac{\alpha}{2}. \end{aligned} \quad (8)$$

Therefore, setting $g_{\lambda, n, \alpha/2} = LCL_G/(\gamma_1 \sigma_{x_0})$ and $g_{\lambda, n, 1-\alpha/2} = UCL_G/(\gamma_1 \sigma_{x_0})$, one can obtain the probability control limits as follows:

$$\begin{aligned} LCL_G &= \sigma_{x_0} \gamma_1 g_{\lambda, n, \alpha/2}, \\ UCL_G &= \sigma_{x_0} \gamma_1 g_{\lambda, n, 1-\alpha/2}, \end{aligned} \quad (9)$$

where $g_{\lambda, n, \alpha}$ is the quantile of distribution of G_Z . Monte carlo numerical strategies should be used to compute $g_{\lambda, n, \alpha}$ since the distribution of the statistic G_i is not available as the quality variable followed by the SN distribution. If $\alpha = 0.0027$, then Eq. (9) can be rewritten as:

$$LCL_R = D_1^* \sigma_{x_0}, \quad UCL_R = D_2^* \sigma_{x_0}, \quad (10)$$

where $D_1^* = \gamma_1 g_{\lambda, n, 0.00135}$ and $D_2^* = \gamma_1 g_{\lambda, n, 0.99865}$. In real cases, for the unknown parameters of the process, the probability control limits of the Gini chart can be obtained by:

$$SN \text{ Gini chart} : \begin{cases} LCL_R = D_3^* \bar{G}, \\ CL_R = \bar{G}, \\ UCL_R = D_4^* \bar{G}, \end{cases} \quad (11)$$

where the charting constants are $D_3^* = D_1^*/(\gamma_1 d_2^*)$ and $D_4^* = D_2^*/(\gamma_1 d_2^*)$. Tables 1 and 2 summarize the numerical charting constants under knowing and unknowing the process parameters in order to facilitate the implementation. We consider different values of λ

Table 1. Charting constants for the known process parameters.

λ	$n = 2$				$n = 3$				$n = 4$				$n = 5$			
	A_1^L	A_1^U	D_1^*	D_2^*	A_1^L	A_1^U	D_1^*	D_2^*	A_1^L	A_1^U	D_1^*	D_2^*	A_1^L	A_1^U	D_1^*	D_2^*
0.0	2.12	2.12	0.00	4.53	1.73	1.73	0.05	3.28	1.50	1.50	0.11	2.82	1.34	1.34	0.20	2.55
0.5	2.11	2.14	0.00	4.24	1.72	1.74	0.05	3.09	1.49	1.51	0.12	2.66	1.34	1.34	0.19	2.39
1.0	2.04	2.22	0.00	3.72	1.68	1.80	0.04	2.72	1.46	1.55	0.10	2.34	1.31	1.37	0.16	2.13
1.5	1.94	2.32	0.00	3.50	1.61	1.87	0.03	2.50	1.41	1.60	0.09	2.19	1.27	1.42	0.14	1.95
2.0	1.85	2.41	0.00	3.27	1.54	1.93	0.03	2.42	0.36	1.65	0.08	2.05	1.23	1.46	0.13	1.85
2.5	1.76	2.48	0.00	3.13	1.48	1.97	0.03	2.31	1.31	1.68	0.08	1.97	1.19	1.49	0.12	1.78
3.0	1.69	2.52	0.00	3.14	1.44	2.00	0.03	2.25	1.27	1.71	0.07	1.92	1.16	1.51	0.12	1.73
3.5	1.63	2.55	0.00	3.07	1.40	2.03	0.03	2.24	1.25	1.72	0.07	1.90	1.14	1.53	0.12	1.72
4.0	1.58	2.57	0.00	3.06	1.36	2.04	0.03	2.24	1.22	1.74	0.07	1.87	1.12	1.54	0.12	1.69

Table 2. Charting constants for the unknown process parameters.

λ	$n = 2$				$n = 3$				$n = 4$				$n = 5$			
	A_2^L	A_2^U	D_3^*	D_4^*	A_2^L	A_2^U	D_3^*	D_4^*	A_2^L	A_2^U	D_3^*	D_4^*	A_2^L	A_2^U	D_3^*	D_4^*
0.0	1.87	1.87	0.00	4.01	1.57	1.57	0.04	3.60	1.20	1.20	0.19	3.10	0.83	0.83	0.19	2.80
0.5	1.86	1.91	0.00	4.05	2.03	2.06	0.06	3.42	1.50	1.55	0.11	2.85	0.84	0.88	0.23	2.63
1.0	1.81	1.99	0.00	4.07	2.01	2.19	0.02	2.89	1.15	1.25	0.11	2.60	1.12	1.19	0.15	2.35
1.5	1.74	2.08	0.00	4.14	1.53	1.78	0.05	2.69	1.20	1.30	0.10	2.36	0.86	0.93	0.18	2.17
2.0	1.62	2.14	0.00	4.19	1.83	2.25	0.02	2.59	0.95	1.22	0.09	2.27	0.72	0.93	0.16	2.05
2.5	1.55	2.21	0.00	4.26	2.20	2.85	0.04	2.44	1.60	2.32	0.07	2.19	1.25	1.59	0.12	1.89
3.0	1.50	2.27	0.00	4.31	2.27	3.12	0.07	2.42	1.40	2.18	0.07	2.13	1.14	1.45	0.12	1.85
3.5	1.46	2.32	0.00	4.34	2.28	3.25	0.04	2.20	1.55	2.43	0.06	2.11	1.33	1.77	0.11	1.81
4.0	1.42	2.35	0.00	4.36	1.19	2.27	0.05	2.30	1.05	1.50	0.06	2.17	1.30	1.90	0.10	1.80

ranging from 0 to 4 to achieve different levels of the skewness coefficient, namely 0.00, 0.02, 0.14, 0.30, 0.45, 0.58, 0.67, 0.73, and 0.78. Moreover, only four sample sizes $n = 2, 3, 4, 5$ are considered since in most practical cases, larger sample sizes are often not available owing to sampling cost. We note that the values in Table 1 and 2 can be used to obtain the charting constants for negative values of λ . Specifically, for $\lambda < 0$ in Table 1, D_1^* and D_2^* are equal to those for $\lambda > 0$ while (A_1^L, A_1^U) is equal to (A_1^U, A_1^L) for $\lambda > 0$ (for example, if $n = 4$ and $\lambda = 0.5$, then $(A_2^L, A_2^U) = (1.49, 1.51)$ and; if $n = 4$ and $\lambda = -0.5$, then $(A_2^L, A_2^U) = (1.51, 1.49)$).

4. Simulation studies

In this section, three simulations as Scenarios I, II, and III are conducted to check the well-behaving of our proposed methodology implemented in statistical software *R* code. The used nomenclatures for mean and variability control charts are as follows:

Mean control charts

$SNG_{\bar{X}}$	SN Gini \bar{X} control chart proposed in Section 3,
$SNL_{\bar{X}}$	SN \bar{X} control chart proposed by Li et al. [1],
$SNT_{\bar{X}}$	SN \bar{X} control chart proposed by Tsai [22],
$CS_{\bar{X}}$	Conventional Shewhart \bar{X} control chart,
$SC_{\bar{X}}$	Skewness correction \bar{X} control chart proposed by Chan and Cui [6],
$WSD_{\bar{X}}$	Weighted standard deviation \bar{X} control chart proposed by Chang and Bai [5].

Variability control charts

SN_G	SN Gini control chart proposed in Section 3,
SN_R	SN R control chart proposed by Li et al. [1],
CS_R	Conventional Shewhart R control chart,
SC_R	Skewness correction R control chart proposed by Chan and Cui [6],
WSD_R	Weighted standard deviation R control chart proposed by Chang and Bai [5].

The adopted scenarios are as follows:

Scenario I: Under the assumption of unknown process parameters, it computes and compares the false alarm rates of mean and variability control charts;

Scenario II: It evaluates the performance of the proposed control charts when the samples are taken from three commonly used skewed models: the Weibull, generalized SN, and skew- t distributions;

Scenario III: It provides a comparison between the mean and variability control charts using power curves.

Without loss of generality, we generate the process measurements from $SN(\lambda)$ and use the nominal value $\alpha = 0.0027$ to design all control charts in the simulation study.

4.1. Scenario I: SN distribution with unknown parameters

Process parameters are usually unknown, and it affects the efficiency in the use of control charts to detect a special cause, since the control limits are usually calculated based on the estimates of these parameters. In the case of unknown process parameters, we first generate data base on the following strategy. Fix n and λ . Under the value of n , generate subgroups of size $m = 50$ from $SN(\lambda)$ to calculate control limits. Then, evaluate the process mean and variability outside these limits and by replication of 5000 times, the false alarm rates, the percentages of process mean and variability outside the control limits are estimated. Table 3 shows the false alarm rates of mean and variability control charts. In Table 3, it is clear that all false alarm rates of \bar{X} control chart, except $CS_{\bar{X}}$, are close to the nominal value. One would expect this fact since all of the control charts used were designed by using the nominal value. For the normal case ($\lambda = 0$), the false alarm rates of all \bar{X} control charts are similar. However, by increasing λ , it can be seen that $SNG_{\bar{X}}$ has the closest values to the nominal value in almost all combinations of n and λ . This shows that $SNG_{\bar{X}}$ performs better than $CS_{\bar{X}}$, $WSD_{\bar{X}}$, $SC_{\bar{X}}$, $SNT_{\bar{X}}$, and $SNL_{\bar{X}}$. Comparing the false alarm rates of SN_G , CS_R , WSD_R , SC_R and SN_R , results depicted in Table 3 show the false alarm rates of WSD_R , SC_R , SN_R , and SN_G are also close to nominal value, and again, with false alarm rates closest to nominal value 0.0027, SN_G outperforms the other competitors.

4.2. Scenario II: Weibull, generalized SN, and skew- t distributions

The second simulation is carried out to check performance of the proposed control charts when the process measurements are generated from the Weibull, skew- t , and generalized SN [26] distributions. In order to achieve four degrees of skewness, $k_3 = 0.0, 0.3, 0.6, 0.9$, the presumed parameter values of the Weibull distribution are 3.6018, 2.6375, 2.0475, and 1.6625 for the shape parameter, respectively, and 1 for the scale one. The shape parameter of skew- t distribution is changed among the values of 0, 0.1145, 0.2368, and

Table 3. False alarm rates of mean and variability control charts in Scenario I.

n	λ	Mean						Variability				
		$SNG_{\bar{X}}$	$SNL_{\bar{X}}$	$SNT_{\bar{X}}$	$CS_{\bar{X}}$	$SC_{\bar{X}}$	$WSD_{\bar{X}}$	$SNG_{\bar{X}}$	SN_R	CS_R	SC_R	WSD_R
2	0	0.00450	0.00450	0.00512	0.00439	0.00450	0.00502	0.00340	0.00340	0.00433	0.00393	0.00350
	1	0.00473	0.00473	0.00545	0.00492	0.00491	0.00533	0.00372	0.00372	0.00462	0.00443	0.00587
	2	0.00462	0.00462	0.00601	0.00643	0.00533	0.00560	0.00460	0.00460	0.00590	0.00460	0.00466
	4	0.00694	0.00694	0.00893	0.00983	0.00702	0.00684	0.00531	0.00531	0.00738	0.00686	0.00829
3	0	0.00383	0.00368	0.00402	0.00360	0.00384	0.00396	0.00315	0.00299	0.00388	0.00365	0.00380
	1	0.00312	0.00366	0.00405	0.00398	0.00388	0.00427	0.00312	0.00357	0.00410	0.00401	0.00373
	2	0.00370	0.00379	0.00391	0.00482	0.00402	0.00420	0.00383	0.00436	0.00525	0.00498	0.00511
	4	0.00334	0.00383	0.00367	0.00670	0.00383	0.00353	0.00455	0.00539	0.00707	0.00680	0.00464
4	0	0.00320	0.00341	0.00365	0.00341	0.00350	0.00373	0.00279	0.00288	0.00343	0.00320	0.00305
	1	0.00369	0.00360	0.00374	0.00378	0.00383	0.00401	0.00310	0.00352	0.00440	0.00391	0.00356
	2	0.00339	0.00374	0.00368	0.00446	0.00380	0.00394	0.00354	0.00412	0.00601	0.00493	0.00393
	4	0.00298	0.00342	0.00371	0.00519	0.00421	0.00298	0.00417	0.00480	0.00724	0.00576	0.00514
5	0	0.00320	0.00330	0.00342	0.00334	0.00338	0.00348	0.00266	0.00284	0.00327	0.00304	0.00292
	1	0.00353	0.00338	0.00347	0.00350	0.00350	0.00372	0.00299	0.00321	0.00410	0.00342	0.00348
	2	0.00319	0.00353	0.00359	0.00442	0.00380	0.00390	0.00317	0.00368	0.00477	0.00401	0.00367
	4	0.00308	0.00369	0.00382	0.00521	0.00422	0.00392	0.00385	0.00510	0.00696	0.00567	0.00530

0.3792 relative to k_3 , while the location, scale, and the degrees of freedom parameters are 0, 1, and 10, respectively. For the generalized SN distribution, the shape parameter is changed among the values of 0.0, 0.1564, 0.2893, and 0.4212 and the location, scale, and the other shape parameters are 0, 1, and 2 to achieve k_3 . In each replication of 5000 trials, samples of size n are generated from all three distributions. Then, the false alarm rates are estimated based on the percentages of process mean and variability according to Gini and R outside the SN Gini and Li's control limits, respectively. Numerical results reported in Table 4 show that the false alarm rates for both mean and variability control charts are far from the nominal value for the Weibull, generalized SN, or skew- t distributions. These results also show that the proposed control charts will have inflated false alarm rates if the skew-normality assumption is violated. Therefore, we would suggest that confirmation of the skew-normality assumption via statistical test is required to construct SN control charts. By comparing the SN Gini control charts with the SN Li's control charts, it can be seen that the false alarm rates of SN Gini control charts

are closer to the nominal value of 0.0027 for most of the combinations of parameters with different sample sizes, reflecting the outperformance of SN Gini control charts versus SN Li's control charts.

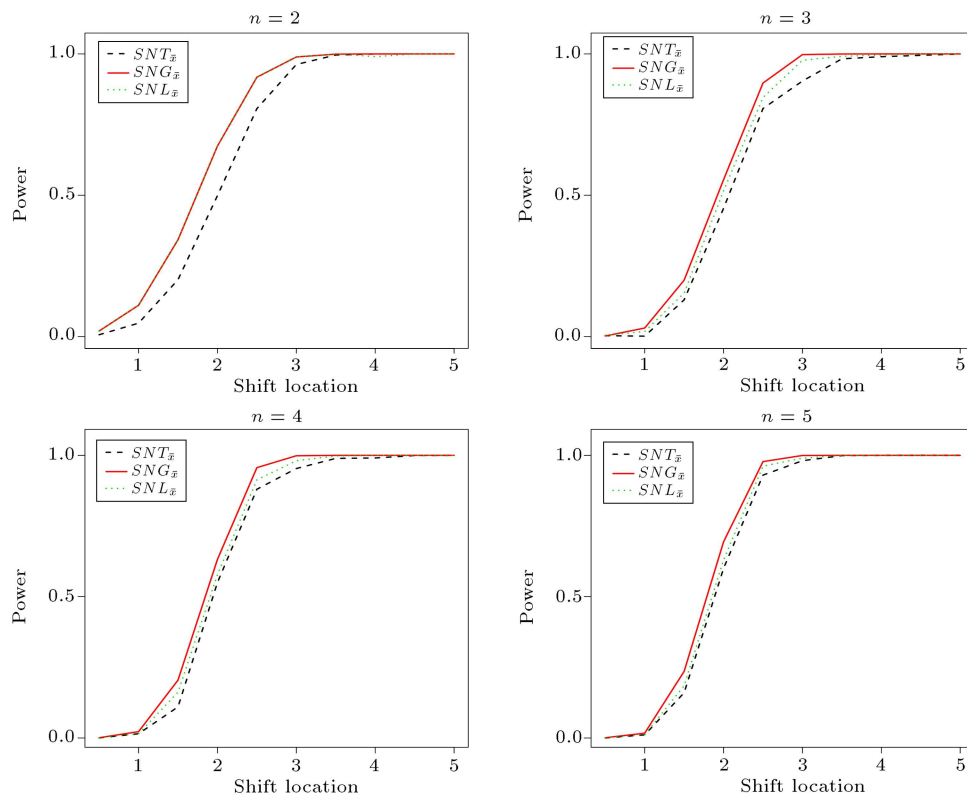
4.3. Scenario III: Power curves of SN distribution

For the comparison purpose, we first exploited the power curves as a performance measure in this section. We compare our proposed $SNG_{\bar{X}}$ chart with those previously introduced by [1] and [22] ($SNL_{\bar{X}}$ and $SNT_{\bar{X}}$ charts), respectively based on R and S . By considering parameters of SN distribution $(\xi, \sigma, \lambda) = (0, 1, 1)$, location parameter is shifted to $\xi + \delta\sigma$. Figure 1 shows the power curves of $SNG_{\bar{X}}$, $SNL_{\bar{X}}$ and $SNT_{\bar{X}}$ charts as a function of shifted location for some values of n and for $\alpha = 0.0027$. It can be seen that the $SNG_{\bar{X}}$ chart can provide a more powerful platform than $SNT_{\bar{X}}$ chart and a good alternate to $SNL_{\bar{X}}$ chart in terms of discriminatory power for detecting shifts in location parameter.

Similarly, by shifting the scale parameter of SN distribution to $\delta\sigma$, the power curves of SN_G chart and

Table 4. False alarm rates of mean and variability control charts in Scenario II.

n	k_3	Mean						Variability					
		Weibull		Generalized SN		Skew- t		Weibull		Generalized SN		Skew- t	
		$SNG_{\bar{X}}$	$SNL_{\bar{X}}$	$SNG_{\bar{X}}$	$SNL_{\bar{X}}$	$SNG_{\bar{X}}$	$SNL_{\bar{X}}$	SN_G	SN_R	SN_G	SN_R	SN_G	SN_R
2	0.0	0.00386	0.00386	0.00494	0.00494	0.01063	0.01063	0.16181	0.16181	0.18340	0.18340	0.17554	0.17554
	0.3	0.00328	0.00328	0.00593	0.00593	0.01458	0.01458	0.17488	0.17488	0.18967	0.18967	0.19648	0.19648
	0.6	0.00441	0.00441	0.00702	0.00702	0.01379	0.01379	0.19922	0.19922	0.19283	0.19283	0.18980	0.18980
	0.9	0.00643	0.00643	0.01031	0.01031	0.07899	0.07899	0.20334	0.20334	0.20307	0.20307	0.17563	0.17563
3	0.0	0.00285	0.00313	0.00413	0.00430	0.00675	0.00745	0.02421	0.03070	0.04119	0.04503	0.04713	0.07032
	0.3	0.00272	0.00308	0.00490	0.00539	0.00688	0.00757	0.02296	0.03067	0.04059	0.04642	0.05392	0.09175
	0.6	0.00270	0.00304	0.00607	0.00648	0.00836	0.00936	0.02993	0.03577	0.05077	0.05210	0.02596	0.04882
	0.9	0.00392	0.00446	0.00777	0.00809	0.00710	0.00790	0.03917	0.04450	0.06117	0.06368	0.02990	0.04567
4	0.0	0.00278	0.00299	0.00382	0.00416	0.00485	0.00617	0.00490	0.00737	0.01508	0.01669	0.01201	0.01772
	0.3	0.00270	0.00270	0.00425	0.00498	0.00440	0.00612	0.00348	0.00632	0.01682	0.01721	0.01370	0.01831
	0.6	0.00272	0.00288	0.00488	0.00527	0.00522	0.00634	0.00566	0.00843	0.01820	0.01892	0.01496	0.01992
	0.9	0.00315	0.00370	0.00572	0.00657	0.00543	0.00701	0.00685	0.01239	0.02185	0.02310	0.00967	0.01635
5	0.0	0.00269	0.00295	0.00299	0.00337	0.00319	0.00531	0.00129	0.00183	0.00309	0.00327	0.00807	0.01159
	0.3	0.00266	0.00280	0.00312	0.00369	0.00371	0.00549	0.00159	0.00197	0.00346	0.00360	0.00765	0.01131
	0.6	0.00265	0.00291	0.00354	0.00412	0.00295	0.00501	0.00260	0.00325	0.00423	0.00465	0.00342	0.00502
	0.9	0.00307	0.00357	0.00369	0.00455	0.00451	0.00657	0.00309	0.00451	0.00402	0.00456	0.00483	0.00782

**Figure 1.** Power curves of $SNG_{\bar{X}}$, $SNL_{\bar{X}}$, and $SNT_{\bar{X}}$ charts in Scenario III.

SN_R chart are drawn in Figure 2. Although the power curves of both charts coincide for $n = 2$, it is clear that for $n > 2$ the power curves of SN_G chart are above those of SN_R chart for given shifts, which shows a better performance for our proposed SN_G chart.

5. A real data example

This section demonstrates the usefulness of this methodology by performing a Principal Components Analysis (PCA) using the data for $m = 4$ variables

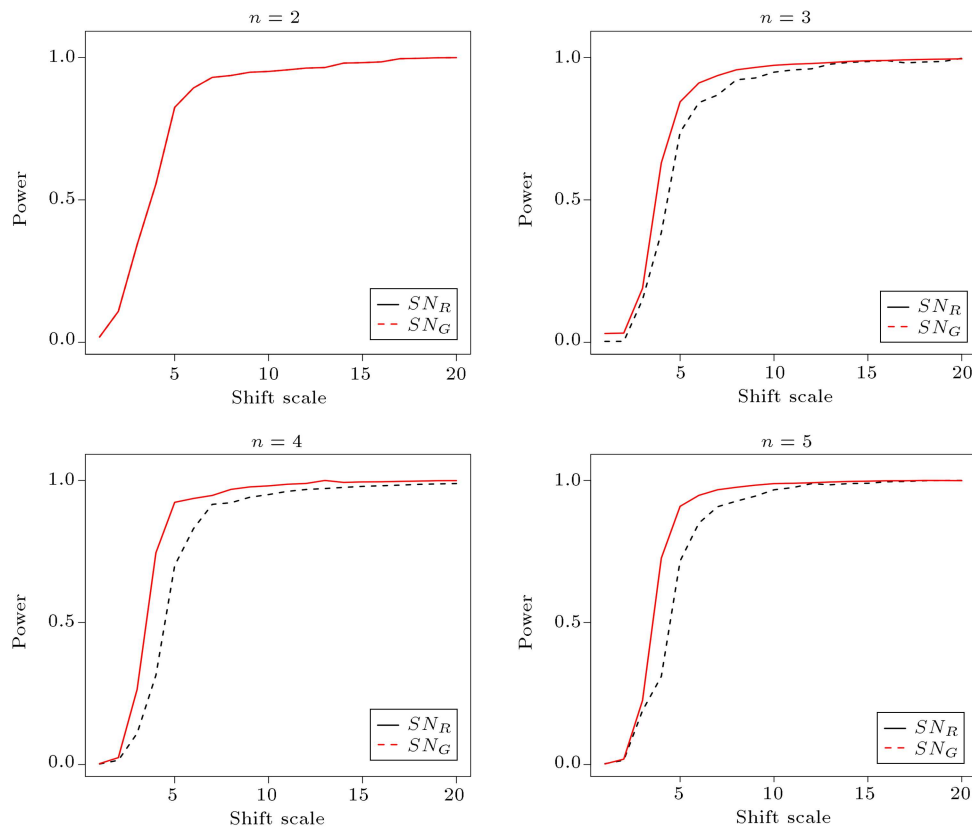


Figure 2. Power curves of SN_R and SN_G charts in Scenario III.

Table 5. Standard chemical process data.

(X_1, X_2, X_3)			(X_1, X_2, X_3)			(X_1, X_2, X_3)		
-0.224	-0.464	-0.662	-0.082	-0.203	0.682	-0.436	0.342	-0.174
0.455	0.715	1.229	0.437	0.237	-0.377	0.669	-0.002	0.230
-0.186	-0.480	-0.907	0.672	0.384	0.903	-0.121	-0.812	-1.279
-0.683	0.093	-0.436	0.063	0.193	-0.028	-1.029	0.503	0.728
-0.080	-0.238	-0.218	0.943	0.272	0.831	0.147	0.655	-0.543
0.217	-0.260	-0.005	-0.611	-0.048	-0.428	0.473	0.066	0.890
-1.130	-1.214	-0.064	1.379	2.443	2.350	0.671	0.534	1.120
1.467	0.945	0.402	1.077	0.949	-0.085	2.126	3.964	3.775
2.237	0.959	0.175	0.362	0.344	0.375	2.119	1.569	1.031
1.147	1.003	1.417	0.667	0.525	1.676	1.944	0.935	-0.130

x_1, x_2, x_3 , and x_4 in [27] (Table 11.6, page 535), which are process variables from a chemical process. For simplicity of illustration, we focus solely on x_1, x_2, x_3 and the standard observations of this data are given in Table 5. Looking at the histogram plot in Figure 3, the skewness of the data is evident. We also apply Shapiro-Wilk goodness-of-fit test to assess the normality of the data. This test results in a p -value of 0.04028, which strongly rejects the normality assumption. Moreover, the goodness-of-fit

test for SN distribution, with the “*GOF*SN” packages in statistical software *R* (*cran.um.ac.ir*), is conducted and it confirms a skew normally distribution of the data with p -value > 0.1 . By fitting the normal and SN distributions to the data and finding the maximum likelihood parameter estimates, the Kolmogorov-Smirnov (KS) test [28] associated with the *PP*-plots of the fitted models is done (Table 6). From the histogram overlaid with the fitted curves of the normal and SN distributions in Figure 3, we can detect that the SN

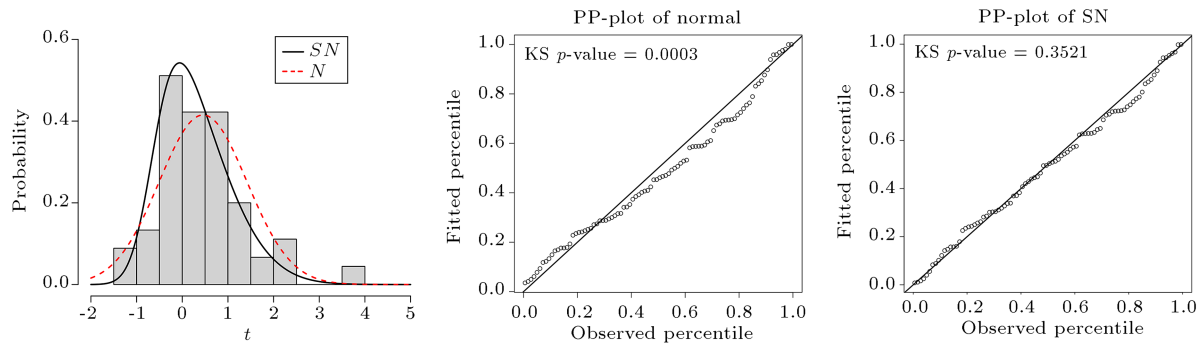


Figure 3. Histogram and PP-plot of standard chemical process data.

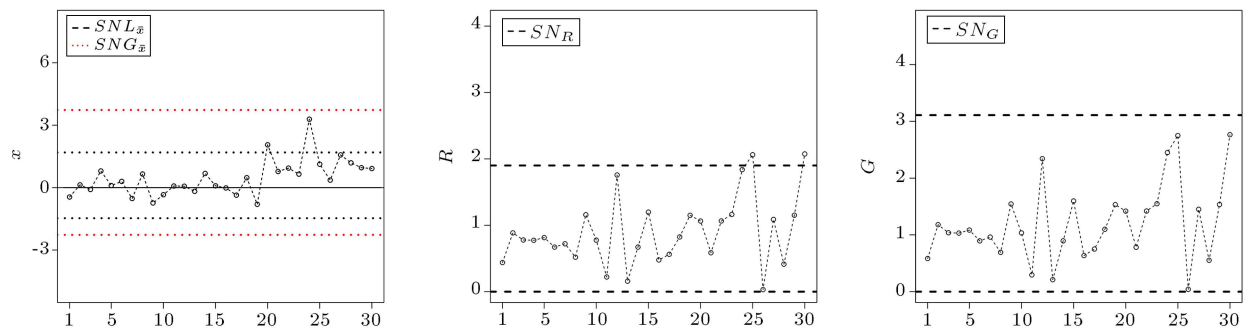


Figure 4. Control chart of standard chemical process data.

Table 6. fitted parameter based on MLE method

Parameter	Normal		SN	
	Estimate	SE	Estimate	SE
μ	0.452	0.044	-0.609	0.029
σ	0.961	0.243	1.431	0.197
λ	—	—	2.991	0.793
$\ell(\theta)$	-131.22		-117.79	
AIC	266.44		241.59	
BIC	271.43		249.09	
KS test	0.722		0.142	
P-value	0.0003		0.3521	

distribution offers a better fit than the normal one to this dataset. The associated p -values of the KS test (0.0003 for the normal distribution and 0.3521 for the SN distribution) and the PP -plots also confirm the SN performance.

We establish $SNL_{\bar{X}}$, $SNG_{\bar{X}}$, SN_R , and SN_G control charts for the SN distribution. We found that $\bar{X} = 0.456$, $\bar{R} = 0.90$, $\bar{G} = 1.21$, and $\hat{\lambda} = 2.91$. Figure 4 (left) depicts graphical plots for $SNL_{\bar{X}}$ and $SNG_{\bar{X}}$ control charts and it can be observed that all points fall within the two considered control limits based on $SNG_{\bar{X}}$ control chart. Figure 4 (center and right) also depicts graphical plots of SN_R and SN_G for both methods by Li et al. [1] and Gini method, with the SN distribution. It can be observed that all points

fall within the two considered control limits, except the two ones that exceed the upper limit of SN_R . This fact reflects that the proposed Gini type SN control limits may not signal a false alarm.

6. Conclusions

This paper proposed Gini control charts to monitor the process mean and variability where the sample was generated from the Skew-Normal (SN) distribution. Some formulae and charting constants aiming to aid and facilitate practical implementation were derived. Numerical results based on the three simulation scenarios and also a real data example showed that the proposed Gini charts could provide considerable improvement over the available methods, especially for highly skewed data. Moreover, it was found that the proposed Gini chart for SN distributed data was superior to control charts based on S and R charts presented by [22] and [1], respectively, in a sense that the discriminatory power of Gini chart was higher than both of them.

Our current approach can be extended to developing the exponentially weighted moving average and cumulative sum control charts for skew normally distributed data. There is a number of avenues for future research. An extension to control chart based on SN distribution appears to be suitable for such situations. Also, another extension is to consider multivariate SN distributions introduced by [23].

References

- Li, C.I., Su, N.C., Su, P.F., et al. "The design of \bar{X} and R control charts for skew-normal distributed data", *Communications in Statistics-Theory and Methods*, **43**(23), pp. 4908–4924 (2014).
- Bittanti, S., Lovera, M., and Moiraghi, L. "Application of non-normal process capability indices to semiconductor quality control", *IEEE Transactions on Semiconductor Manufacturing*, **11**(2), pp. 296–303 (1998).
- Choobineh, F. and Ballard, J.L. "Control-limit of QC charts for skewed distributions using weighted variance", *IEEE Transactions on Reliability*, **36**(4), pp. 473–477 (1987).
- Jacobs, D.C. "Watch out for non-normal distributions", *Chemical Engineering Progress*, **86**, pp. 19–27 (1999).
- Chang, Y.S. and Bai, D.S. "Control charts for positively skewed populations with weighted standard deviations", *Quality and Reliability Engineering International*, **17**(5), pp. 397–406 (2001).
- Chan, L.K. and Cui, H.J. "Skewness correction \bar{X} and R charts for skewed distributions", *Naval Research Logistics*, **50**(6), pp. 555–573 (2003).
- Riaz, M. and Saghir, A. "Monitoring process variability using Gini's mean difference", *Quality Technology and Quantitative Management*, **4**(4), pp. 439–454 (2007).
- Riaz, M. and Abbasi, S.A. "Gini's mean difference based time-varying EWMA charts", *Economic Quality Control*, **24**(2), pp. 269–276 (2009).
- Abbasi, S.A. and Miller, A. "On proper choice of variability control chart for normal and non-normal processes", *Quality and Reliability Engineering International*, **28**(3), pp. 279–296 (2012).
- Schoonhoven, M. and Does, J.M.M. "The \bar{X} control chart under non-normality", *Quality and Reliability Engineering International*, **26**(2), pp. 167–176 (2010).
- Schoonhoven, M., Riaz, M., and Does, J.M.M. "Design and analysis of control charts for standard deviation with estimated parameters", *Journal of Quality Technology*, **43**(4), pp. 307–333 (2011).
- Saghir, A., Saleem, A., and Anjum, M.S. "Use of Gini's mean difference in analysis of means", *Pakistan Journal of Science*, **64**(4), pp. 363–369 (2012).
- Saghir, A. and Lin, Z. "A study on the robustness of G-chart to non-normality", *Communications in Statistics-Simulation and Computations*, **43**(10), pp. 2241–2251 (2014).
- Saghir, A. and Lin, Z. "Designing of Gini-chart for exponential, t , logistic and Laplace distributions", *Communications in Statistics-Simulation and Computations*, **44**(9), pp. 2387–2409 (2015).
- Sindhumol, M.R., Srinivasan, M.R., and Gallo, M. "Robust control charts based on modified trimmed standard deviation and Gini's mean difference", *Journal of Applied Quantitative Methods*, **11**(3), pp. 18–30 (2016).
- Mangold, B. and Konopik, J. "A general class of entropy based control charts (No. 04/2017)", FAU Discussion Papers in Economics (2017). <https://www.iwf.rw.fau.de/research/iwf-discussion-paper-series/>
- Rizwan Iqbal, M. and Hassan, S. "Robust skewness dispersion control charts", *International Journal of Quality and Reliability Management*, **35**(10), pp. 2136–2156 (2018).
- Mehmood, R., Qazi, M.S., and Riaz, M. "On the performance of \bar{X} control chart for known and unknown parameters supplemented with runs rules under different probability distributions", *Journal of Statistical Computation and Simulation*, **88**(4), pp. 675–711 (2018).
- Saeed, N. and Kamal, S. "New EWMA control charts for monitoring mean under Non-normal processes using repetitive sampling", *Iranian Journal of Science and Technology, Transactions A: Science*, **43**(3), pp. 1215–1225 (2019).
- Karagrigoriou, A., Makrides, A., Vonta, I. "On a control chart for the Gini index with simulations", *Communications in Statistics-Simulation and Computation*, **48**(4), pp. 1121–1137 (2019).
- Azzalini, A. "A class of distributions which includes the normal ones", *Scandinavian Journal of Statistics*, **12**, pp. 171–178 (1985).
- Tsai, T.R. "Skew-normal distribution and the design of control charts for averages", *International Journal of Reliability, Quality and Safety Engineering*, **14**(01), pp. 49–63 (2007).
- Azzalini, A. and Capitanio, A. "Statistical applications of the multivariate skew-normal distribution", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(3), pp. 579–602 (1999).
- Gupta, A.K. and Chen, T. "On the sample characterization criterion for normal distributions", *Journal of Statistical Computation and Simulation*, **73**, pp. 155–163 (2003).
- Riaz, M. "A dispersion control chart", *Communications in Statistics-Simulation and Computation*, **37**(6), pp. 1239–1261 (2008).
- Arellano-Valle, R.B., Gómez, H.W., and Quintana, F.A. "A new class of skew normal distributions", *Communication in Statistics-Theory and Methods*, **33**(7), pp. 1465–1480 (2003).
- Montgomery, D.C., *Introduction to Statistical Quality Control*, John Wiley & Sons. Inc., USA (1997).
- Smirnov, N. "Table for estimating the goodness of fit of empirical distributions", *The Annals of Mathematical Statistics*, **19**(2), pp. 279–281 (1948).

Biographies

Mahsa Sasaei earned her BSc and MSc degrees from Shahid Bahonar University of Kerman (SBUK), Iran, both in Statistics. At present, she is a PhD student at SBUK and has published several journal and conference papers in various fields. Her research interests include fuzzy statistics, statistical quality control, and distributions theory.

Reza Pourmousa is an Assistant Professor at SBUK. He received his BSc and MSc degrees in Statistics from Esfahan University and Shahid Beheshti University and PhD degree at SBUK. His research interests include

fuzzy statistics, statistical modeling, and distributions theory.

Mashaallah Mashinchi received his BSc and MSc in Statistics from Ferdowsi University of Mashhad and Shiraz University, Iran, respectively, and his PhD in Mathematics from Waseda University, Japan. He is an Emeritus Professor at the Department of Statistics at SBUK. He is the founding editor and a member of the editorial board of the Iranian Journal of Fuzzy Systems as well as some other esteemed journals. His current interest is in fuzzy mathematics, especially in statistics, statistical quality control, decision-making, and algebraic systems.