



Economic Statistical Design of X Bar Control Chart for Non-Normal Symmetric Distribution of Quality Characteristic

Kristina Veljkovic^a, Halima Elfaghihe^a, Vesna Jevremovic^a

^aDepartment of Probability and Statistics, Faculty of Mathematics, University of Belgrade, Studentski trg 16, Belgrade, Serbia

Abstract. In economic statistical design of a control chart, the economic-loss function is minimized subject to a constrained minimum value of power, maximum value of probability of false alarms and average time to signal an expected shift. This paper is concerned with the optimum economic statistical design of the X bar chart when quality characteristic has non-normal symmetric distribution. We considered three types of distributions: Student distribution, standard Laplace distribution and logistic distribution. For each of these distributions, we calculated theoretical distribution of standardized sample mean (or its best approximation) and approximated it with normal, Pearson VII and Johnson SU distributions. For considered example, constrained minimization of expected loss function was done using genetic algorithm in statistical software R. We compared results of economic statistical design of X-bar chart for theoretical distribution of standardized sample mean with the results for normal, Pearson and Johnson distributions. We found that, for all chosen distributions of quality characteristic, Pearson VII distribution and Johnson SU distribution give results very close to results based on theoretical distribution of standardized sample mean, while normal distribution gives much worse fit.

1. Introduction

The X bar chart is extensively used in practice to monitor a change in the process mean. It is usually assumed that quality characteristic under the surveillance of an X-chart has a normal distribution. On the other hand, occurrence of non-normal data in industry is quite common (see Alloway and Raghavachari, 1991; Janacek and Meikle, 1997). Violation of normality assumption results in incorrect control limits of control charts (Alwan, 1995). Mislplaced control limits lead to inappropriate charts that will either fail to detect real changes in the process or which will generate spurious warnings when the process has not changed.

In the case of non-normal symmetric distribution of quality characteristics, no recommendations, except the use of normal distribution, are given in the quality control literature. Approximation of the distribution of sample mean with normal distribution is based on the central limit theorem, but in practice small sample sizes are usually used.

We will consider three types of non-normal symmetric distributions of quality characteristic: Student distribution, Laplace distribution and logistic distribution. These distributions are chosen because of

2010 *Mathematics Subject Classification.* 60E05; 62E17; 62P30; 90C59

Keywords. X-bar control chart, economic statistical design, non-normal symmetric distribution, Pearson type VII distribution, Johnson SU distribution

Received: 01 September 2014; Accepted: 07 January 2015

Communicated by Dragana Cvetković-Ilić

Email address: kristina@matf.bg.ac.rs (Kristina Veljkovic)

their applications in various disciplines (economics, finance, engineering, hydrology, etc, see for instance Ahsanullah et al., 2014; Balakrishnan, 1991; Kotz et al., 2001). For each of these distributions, we calculated theoretical distribution of standardized sample mean (or its best approximation) and then we approximated it with normal, Pearson VII and Johnson SU distributions. Pearson system of distributions, Johnson's transformations and normal distribution are known to provide approximations to a wide variety of observed distributions (Johnson et al., 1994).

In this paper, we will be concerned with the optimum economic statistical design of the X-bar chart for chosen distributions of quality characteristic. In economic statistical design of a control chart, the economic-loss function is minimized subject to a constrained minimum value of power and maximum value of the Type I error probability and ATS (average time to signal) an expected shift (Saniga, 1989). Constrained minimization of expected loss function was done in statistical software R, using genetic algorithm *ga* from R package GA (Scrucca, 2013). Results of economic statistical design of X-bar chart for theoretical distribution of standardized sample mean and normal, Pearson VII and Johnson SU approximations are compared, with results based on the theoretical distribution of sample mean as the 'golden' standard.

Our objective is, using previous theoretical results concerning distributions of standardized sample mean, to provide a reliable method for design of X bar control chart that takes into account economic factors while achieving desirable statistical properties in the case when distribution of the data is symmetric, but non-normal.

The rest of this paper is organized as follows: basic assumptions about the process and methods of design of X-bar control chart are discussed in Section 2. Description of distribution of quality characteristic and distribution of standardized sample mean is given in Sections 3 and 4, respectively. Normal, Pearson and Johnson approximations are considered in Section 5. An illustrative example on the construction of the economic statistical design of X bar chart for theoretical distribution of standardized sample mean and its normal, Pearson VII and Johnson SU approximations is given in Section 6. Conclusions are drawn in Section 7.

2. Economic statistical design of the X bar chart

2.1. Basic assumptions

We shall state briefly the basic assumptions about the process. It is presumed that a process begins in in-control state with mean μ_0 and that single assignable cause of magnitude δ results in a shift in the process mean from μ_0 to either $\mu_0 - \delta\sigma$ or $\mu_0 + \delta\sigma$, where σ is a standard deviation which is assumed to remain stable.

The process is monitored by the X bar chart with central line μ_0 and upper and lower control limits, respectively, $\mu_0 + k\sigma/\sqrt{n}$ and $\mu_0 - k\sigma/\sqrt{n}$, where n represents the sample size and k width of control limits. Samples are taken from the process every h hours and the sample mean is plotted on the X-bar chart. If a sample mean exceeds control limits, it is assumed that some shift in the process mean has occurred and a search for the assignable cause is initiated. The process is allowed to continue operating during the search for the assignable cause. The parameters μ_0 , δ and σ are assumed to be known.

The assignable cause is assumed to occur according to a Poisson process with an intensity of λ occurrences per hour. That is, assuming that the process begins in the in-control state, the time interval that the process remains in control is an exponential random variable with mean $\frac{1}{\lambda}h$. Therefore, given the occurrence of the assignable cause in the interval between the j th and $(j+1)$ st sample, the expected time of occurrence within the interval between samples is

$$\tau = \frac{\int_{jh}^{(j+1)h} (t - jh)\lambda e^{-\lambda t} dt}{\int_{jh}^{(j+1)h} \lambda e^{-\lambda t} dt} = \frac{1 - (1 + \lambda h)e^{-\lambda h}}{\lambda(1 - e^{-\lambda h})}.$$

2.2. Design parameters

Various kinds of parameters are considered in the economic statistical design of the control chart and they are denoted by following symbols:

- n : Sample size.
- h : Sampling interval.
- k : Width of control limits.
- σ : Standard deviation of observations.
- μ_0 : In-control process mean.
- μ_1 : Out-of control process mean.
- δ : Shifts in process mean in standard deviation units when assignable cause occurs ($\delta = |\mu_1 - \mu_0|/\sigma$).
- λ : It is assumed that in-control time follows an exponential distribution with mean $\frac{1}{\lambda}$.
- τ : Expected time of occurrence of the assignable cause, given it occurs between the j -th and $(j + 1)$ -st samples.
- α : Probability of type I error.
- β : Probability of type II error.
- ATS: Average time to signal an expected shift, $ATS = \frac{h}{1-\beta}$.
- g : The time required to take a sample and interpret the results.
- D : The time required to find an assignable cause following an out-of-control signal.
- a_1 : Fixed cost per sample.
- a_2 : Variable cost per unit sampled.
- a_3 : Cost for searching and repairing the assignable cause.
- a'_3 : Cost of investigating a false alarm.
- V_0 : Profit per hour earned by the process operating in-control.
- V_1 : Profit per hour earned by the process operating out-of-control ($V_0 > V_1$).
- a_4 : Hourly penalty cost associated with production in the out-of-control state, $a_4 = V_0 - V_1$
- κ : Expected number of false alarms, $\kappa = \alpha \frac{e^{-\lambda h}}{1-e^{-\lambda h}}$.
- $E(T)$: Expected length of a production cycle. The production cycle time T is defined as the interval of time from the start of production (the process is assumed to start in the in-control state), following an adjustment to the detection and elimination of the assignable cause.
- $E(C)$: Expected cost per cycle. Total cost C consists of the costs of testing and sampling, the costs associated with investigating an out-of-control signal and with the repair or correction of any assignable causes found and the costs associated with the production of nonconforming items.
- s : Expected number of samples taken within a cycle, $s = \frac{E(T)}{h}$.

2.3. The expected cycle time

Each production cycle begins with the process in the in-control state and continues until process monitoring via control charts results in an out-of-control signal. Following an adjustment in which the process is returned to the in-control state, a new cycle begins.

Length of a production cycle T is partitioned into five time intervals, $I_a, I_b, I_c, I_d,$ and I_e (Figure 1).

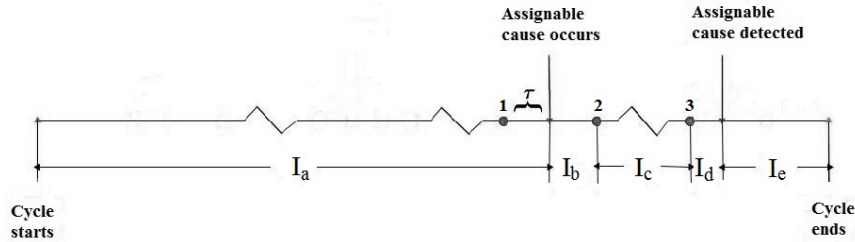


Figure 1: Intervals of production cycle

- I_a : The time until the assignable cause occurs from the start of the cycle.
- I_b : The time until the next sample is taken from the end of I_a .
- I_c : The time until the chart gives an out-of-control signal from the end of I_b .
- I_d : The time to analyze the sample and chart the result from the end of I_c .
- I_e : The time to discover the assignable cause and repair the process from the end of I_d .

Three time points are denoted on Figure 1: point 1 represents last sample before the occurrence of the assignable cause, point 2 first sample taken after assignable cause and point 3 the moment when lack of control is detected. Process remains in-control state during time interval I_a and it is in out-control state during intervals I_b, I_c, I_d and I_e .

Expected lengths of each of these time intervals are, respectively, equal to

$$E(I_a) = \frac{1}{\lambda}, E(I_b) = h - \tau, E(I_c) = ATS - h, E(I_d) = gn, E(I_e) = D.$$

Then, expected length of a cycle is

$$E(T) = \frac{1}{\lambda} + \frac{h}{1 - \beta} - \tau + gn + D.$$

2.4. The expected cost

Expected cost during a production cycle is equal to

$$E(C) = V_0 \cdot \frac{1}{\lambda} + V_1 \cdot \left(\frac{h}{1 - \beta} - \tau + gn + D \right) - a_3 - a'_3 \kappa - (a_1 + a_2 n)s$$

2.5. The expected loss per hour

The expected loss function is equal to

$$E(L) = V_0 - \frac{E(C)}{E(T)} = \frac{a_1 + a_2}{h} + \frac{a_4 \left(\frac{h}{1-\beta} - \tau + gn + D \right) + a_3 + a'_3 \kappa}{\frac{1}{\lambda} + \frac{h}{1-\beta} - \tau + gn + D}.$$

2.6. Methods of design

The use of control chart requires one to select a sample size, a sampling frequency and control limits for the chart. Selection of these three parameters is usually called the design of the control chart. Currently, there are three general methods for design of control charts: statistical, economic and economic statistical design. In statistically designed control charts, control-limit width (which determines the Type I error probability) and power are preselected. Then, sample size and sampling frequency (if ATS is specified) can be determined (Woodall, 1985).

The objective of economic design is to find the sample size, control-limit width, and sampling frequency that minimize the loss in profit composed of several given costs. Duncan (1956) first proposed an economic model for the optimum economical design of the X-bar control chart. This research stimulated many subsequent studies on the subject. Lorenzen and Vance (1986) proposed a general method for determining the economic design parameters of control charts.

The weakness of statistically designed control charts are high operating costs. On the other hand, economic design does not consider the probability of type I and type II errors which can lead to high probability of false alarms and low power. Saniga (1989) proposed economic statistical model which combines properties of economic and statistical design: it is a method for economic design of control charts that have bounds on type I and type II error probabilities and the average time to signal (ATS) an expected shift. Economic statistical design considers economic factors while achieving desirable statistical properties and, therefore, represents the improvement of economic or statistical design.

2.7. Economic statistical model

The economic statistical design of a control chart is defined as a design in which the expected economic-loss cost function $E(L)$ is minimized subject to a constrained minimum value of power and maximum value of the Type I error probability and ATS an expected shift. We wish to find sample size, control-limit width, and sampling frequency to

$$\begin{aligned} & \text{minimize } E(L) \\ & \text{subject to} \\ & \alpha \leq \alpha_0, p \geq p_0, ATS \leq ATS_0, \end{aligned} \tag{1}$$

where $p = 1 - \beta$, α_0 , p_0 and ATS_0 are, respectively, power and the desired bounds on the Type I error probability, power, and ATS an expected shift.

3. Distribution of quality characteristic

In this paper we consider three types of non-normal symmetric distributions of quality characteristic X : Student's distribution, Laplace (double exponential) and logistic distribution (Đorić et al. 2007, Johnson et al. 1994, Johnson et al. 1995). All chosen distributions have heavier tails than normal distribution.

Distributions are given by their density functions. We will use following notations for mean, variance, skewness and kurtosis of quality characteristic, respectively: $\mu = E(X)$, $\sigma^2 = Var(X)$, $\alpha_3 = \frac{E(X-E(X))^3}{\sigma^{\frac{3}{2}}}$, $\alpha_4 = \frac{E(X-E(X))^4}{\sigma^4}$. As all chosen distributions are symmetric around the mean, $\alpha_3 = 0$.

1. Student's distribution $t(\nu, \mu, \eta)$

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\eta \sqrt{\pi\nu} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{1}{\nu} \left(\frac{x-\mu}{\eta}\right)^2\right)^{-\frac{\nu+1}{2}}, x \in \mathbb{R},$$

where ν ($\nu \in \mathbb{R}$) is the shape parameter, η ($\eta > 0$) is scale parameter and μ ($\mu \in \mathbb{R}$) location parameter. Variance and kurtosis are, respectively, equal to

$$\sigma^2 = \frac{\nu}{\nu-2} \eta^2, \nu > 2, \quad \alpha_4 = 3 + \frac{6}{\nu-4}, \nu > 4.$$

Random variable $Y = \frac{X-\mu}{\eta}$ has Student distribution $t(\nu)$ with density function

$$f(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}, y \in \mathbb{R}.$$

2. Laplace distribution $L_2(\eta, \mu)$

$$f(x) = \frac{1}{2\eta} \exp\left\{-\frac{|x-\mu|}{\eta}\right\}, x \in \mathbb{R},$$

where μ ($\mu \in \mathbb{R}$) is location parameter and η ($\eta > 0$) scale parameter. Variance and kurtosis are, respectively, equal to

$$\sigma^2 = 2\eta^2, \quad \alpha_4 = 6.$$

Random variable $Y = X - \mu$ has standard Laplace distribution $L_1(\eta)$ with density function

$$f_Y(y) = \frac{1}{2\eta} \exp\left\{-\frac{|y|}{\eta}\right\}, y \in \mathbb{R}.$$

3. Logistic distribution $LGS(\mu, \eta)$

$$f(x) = \frac{\exp\left\{-\frac{x-\mu}{\eta}\right\}}{\eta \cdot \left(1 + \exp\left\{-\frac{x-\mu}{\eta}\right\}\right)^2}, x \in \mathbb{R},$$

where μ ($\mu \in \mathbb{R}$) is the location parameter and η ($\eta > 0$) scale parameter. Variance and kurtosis are, respectively, equal to

$$\sigma^2 = \frac{\eta^2 \pi^2}{3}, \quad \alpha_4 = 4.2.$$

Random variable $Y = \frac{X-\mu}{\eta}$ has standard logistic distribution with density function

$$f_Y(y) = \frac{e^{-y}}{(1 + e^{-y})^2}, y \in \mathbb{R},$$

4. Distribution of standardized sample mean

4.1. Sample from Student's distribution

Witkovský (2001, 2004) proposed a method for numerical evaluation of the distribution function and the density function of a linear combination of independent Student's t random variables. The method is based on the inversion formula which leads to the one-dimensional numerical integration.

Let (X_1, X_2, \dots, X_n) be a sample from Student's $t(\nu, \mu, \lambda)$ distribution. Then $Y_k = \frac{X_k - \mu}{\eta}$, $k = 1, 2, \dots, n$ are independent Student's t_ν random variables with ν degrees of freedom. Further, let $Y = \sum_{k=1}^n Y_k$ be sum of these variables and let $\phi_{Y_k}(t)$ denote the characteristic function of Y_k . The characteristic function of Y is

$$\phi_Y(t) = \prod_{k=1}^n \phi_{Y_k}(t),$$

where

$$\phi_{Y_k}(t) = \frac{1}{2^{\frac{\nu}{2}-1} \Gamma(\frac{\nu}{2})} \left(\nu^{\frac{1}{2}} |t| \right)^{\frac{\nu}{2}} K_{\nu/2} \left(\nu^{\frac{1}{2}} |t| \right),$$

where $K_\alpha(z)$ denotes modified Bessel function of the second kind.

For more details see Witkovský (2001). Note that the characteristic function of the Student's random variable is a real function.

The cumulative distribution function (cdf) $F_Y(y)$ of random variable Y is according to the inversion formula due to Gil-Pelaez (1951) given by

$$\begin{aligned} F_Y(y) &= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \operatorname{Im} \left(\frac{e^{-ity} \phi_Y(t)}{t} \right) dt = \\ &= \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin(ty) \phi_Y(t)}{t} dt \end{aligned} \quad (2)$$

and the probability density function (pdf) $f_Y(y)$ of Y is given by

$$\begin{aligned} f(y) &= \frac{1}{\pi} \int_0^\infty \operatorname{Re}(e^{-ity} \phi_Y(t)) dt = \\ &= \frac{1}{\pi} \int_0^\infty \cos(ty) \phi_Y(t) dt \end{aligned} \quad (3)$$

For any chosen y algorithm *tdist* in R package *tdist* (Witkovsky and Savin, 2005) evaluates the integrals in (2) and (3) by multiple p -points Gaussian quadrature over the real interval $t \in (0, 10\pi)$. The whole interval is divided into m subintervals given by pre-specified limits and the integration over each subinterval is done with p -points Gaussian quadrature which involves base points b_{ij} , and weight factors w_{ij} , $i = 1, 2, \dots, p$, $j = 1, 2, \dots, m$. So,

$$\begin{aligned} F_Y(y) &\approx \frac{1}{2} + \frac{1}{\pi} \sum_{j=1}^m \sum_{i=1}^p \frac{\sin(b_{ij}y)}{b_{ij}} w_{ij} \phi_Y(b_{ij}) = \\ &= \frac{1}{2} + \frac{1}{\pi} \sum_{j=1}^m \sum_{i=1}^p \frac{\sin(b_{ij}y)}{b_{ij}} W_{ij}, \\ f_Y(y) &\approx \frac{1}{\pi} \sum_{j=1}^m \sum_{i=1}^p \cos(b_{ij}y) w_{ij} \phi_Y(b_{ij}) = \\ &= \frac{1}{\pi} \sum_{j=1}^m \sum_{i=1}^p \cos(b_{ij}y) W_{ij}, \end{aligned}$$

where $W_{ij} = w_{ij}\phi_Y(b_{ij})$. For evaluation of $F_Y(y)$ and $f_Y(y)$ in many different points the algorithm requires only one evaluation of the weights W_{ij} , for $i = 1, 2, \dots, p$ and $j = 1, 2, \dots, m$, which directly depend on the characteristic function $\phi_Y(\cdot)$ and do not depend on y .

Then, distribution function and density function of standardized sample mean $T_n = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} = Y \sqrt{\frac{v-2}{nv}}$ are equal to

$$F_{T_n}(t) = P\left\{Y \sqrt{\frac{v-2}{nv}} \leq t\right\} = F_Y\left(\sqrt{\frac{nv}{v-2}}t\right)$$

$$f_{T_n}(t) = \sqrt{\frac{nv}{v-2}} \cdot f_Y\left(\sqrt{\frac{nv}{v-2}}t\right), t \in \mathbb{R}$$

Skewness and kurtosis of standardized sample mean are, respectively, equal

$$\alpha_{3,T_n} = 0, \quad \alpha_{4,T_n} = 3 + \frac{6}{n(v-4)}.$$

Probability of I type error and power are, respectively, equal to

$$\alpha = 2F_{T_n}(-k), \quad 1 - \beta = F_{T_n}(-k - \delta \sqrt{n}) + F_{T_n}(-k + \delta \sqrt{n}).$$

4.2. Sample from Laplace distribution

Let (X_1, X_2, \dots, X_n) be a sample from Laplace distribution $L_2(\mu, \eta)$. Then, $Y_k = X_k - \mu$, $k = 1, 2, \dots, n$ are independent random variables with standard Laplace distribution $L_1(\eta)$.

We will use the property that if two independent random variables have exponential $\varepsilon(\eta)$ distribution, then their difference has standard Laplace distribution. Further, standard exponential distribution is gamma distribution, $\Gamma(1, \eta)$. Sum of n independent variables with $\Gamma(1, \eta)$ distribution is gamma distribution $\Gamma(n, \eta)$. In that way, we conclude that sum of n independent random variables Y_1, Y_2, \dots, Y_n with standard Laplace distribution can be written as the difference of two random variables with gamma distribution $\Gamma(n, \eta)$ which is called bilateral gamma distribution.

Bilateral gamma distribution is symmetric around 0, with density function for $y > 0$ (Küchler and Tappe, 2008)

$$f(y) = \left(\frac{\eta}{2}\right)^n \cdot \frac{1}{(n-1)!} \sum_{k=0}^{n-1} a_k y^k e^{-\eta y},$$

where the coefficients $(a_k)_{k=0, \dots, n-1}$ are given by

$$a_k = \binom{n-1}{k} \frac{1}{(2\eta)^{n-1-k}} \prod_{l=0}^{n-2-k} (n+l), \quad a_{n-1} = 1.$$

Probability distribution function, for $y > 0$ is

$$F_Y(y) = \frac{1}{2} + \left(\frac{\eta}{2}\right)^n \cdot \frac{1}{(n-1)!} \sum_{k=0}^n \frac{a_k}{\eta^{k+1}} \gamma(k+1, \eta x),$$

where $\gamma(n, x)$ is lower incomplete gamma function.

For evaluation of cdf we will use following recursive relation for incomplete gamma function

$$\gamma(n, x) = (n-1)\gamma(n-1, x) - x^{n-1}e^{-x}$$

and $\gamma(1, x) = 1 - e^{-x}$.

Then, distribution function and density function of standardized sample mean $T_n = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} = \frac{Y}{\eta \cdot \sqrt{2n}}$

$$F_{T_n}(t) = P\left\{\frac{Y}{\eta \cdot \sqrt{2n}} \leq t\right\} = F_Y(\eta \sqrt{2nt}),$$

$$f_{T_n}(t) = \eta \sqrt{2n} \cdot f_Y(\eta \sqrt{2nt}), t \in \mathbb{R}$$

Skewness and kurtosis of standardized sample mean are, respectively, equal

$$\alpha_{3,T_n} = 0, \quad \alpha_{4,T_n} = 3 + \frac{3}{n}.$$

Probability of I type error and power are, respectively equal to

$$\alpha = 2F_{T_n}(-k), \quad 1 - \beta = F_{T_n}(-k - \delta \sqrt{n}) + F_{T_n}(-k + \delta \sqrt{n}).$$

4.3. Sample from logistic distribution

Let (X_1, X_2, \dots, X_n) be a random sample from logistic distribution $LGS(\mu, \eta)$.

George and Mudholkar (1983) showed that a standardized Student's t-distribution provides a very good approximation for the distribution of a convolution of n i.i.d. logistic variables. These authors then compared three approximations: (1) standard normal approximation, (2) Edgeworth series approximation correct to order n^{-1} , and (3) Student's t approximation with $\nu = 5n + 4$ degrees of freedom. They showed that the third provides a very good approximation.

Gupta and Han (1992) considered the Edgeworth series expansions up to order n^{-3} for the distribution of the standardized sample mean

$$T_n = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}.$$

Distribution function is given by

$$F_{T_n}(t) \approx \Phi(t) - \varphi(t) \left(\frac{1}{n} \left(\frac{1}{4!} \cdot \frac{6}{5} H_3(t) \right) + \frac{1}{n^2} \left(\frac{1}{6!} \cdot \frac{48}{7} H_5(t) + \frac{35}{8!} \left(\frac{6}{5} \right)^2 H_7(t) \right) \right) + \frac{1}{n^3} \left(\frac{1}{8!} \cdot \frac{432}{5} H_7(t) + \frac{210}{10!} \frac{48}{7} \cdot \frac{6}{5} H_9(t) + \frac{5775}{12!} \left(\frac{6}{5} \right)^3 H_{11}(t) \right), t \in \mathbb{R},$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are standard normal pdf and cdf and $H_j(x)$ is the Hermite polynomial.

Then, probability density function of standardized sample mean is

$$f_{T_n}(t) \approx \varphi(t) \left(1 + \frac{1}{n} \cdot \frac{1}{4!} \cdot \frac{6}{5} \cdot (tH_3(t) - H'_3(t)) + \frac{1}{n^2} \left(\frac{1}{6!} \cdot \frac{48}{7} \cdot (tH_5(t) - H'_5(t)) + \frac{35}{8!} \left(\frac{6}{5} \right)^2 (tH_7(t) - H'_7(t)) \right) + \frac{1}{n^3} \left(\frac{1}{8!} \cdot \frac{432}{5} \cdot (tH_7(t) - H'_7(t)) + \frac{210}{10!} \frac{48}{7} \cdot \frac{6}{5} (tH_9(t) - H'_9(t)) + \frac{5775}{12!} \left(\frac{6}{5} \right)^3 (tH_{11}(t) - H'_{11}(t)) \right) \right), t \in \mathbb{R}$$

Gupta and Han (1992) compared this approximation with the approximations mentioned earlier, and they showed the approximation to be far better than even the Student's t-approximation suggested by George and Mudholkar (1983).

Skewness and kurtosis of standardized sample mean are, respectively, equal

$$\alpha_{3,T_n} = 0, \quad \alpha_{4,T_n} = 3 + \frac{1.2}{n}.$$

Probability of I type error and power are, respectively equal to

$$\alpha = 2F_{T_n}(-k), \quad 1 - \beta = F_{T_n}(-k - \delta \sqrt{n}) + F_{T_n}(-k + \delta \sqrt{n}).$$

5. Approximations of the distribution of sample mean

We will approximate distribution of standardized sample mean T_n with normal distribution, Pearson's type VII distribution and Johnson's SU distribution.

5.1. Normal distribution

Probability of I type error and power are, respectively equal to

$$\alpha = 2\Phi(-k), \quad 1 - \beta = \Phi(-k - \delta \sqrt{n}) + \Phi(-k + \delta \sqrt{n}),$$

where $\Phi(\cdot)$ is a probability distribution function of standard normal distribution.

5.2. Pearson type VII distribution

Probability density function is equal to (Johnson et al. 1994, Đorić et al. 2007)

$$f(x) = \frac{1}{\alpha B\left(m - \frac{1}{2}, \frac{1}{2}\right)} \cdot \left(1 + \frac{x^2}{\alpha^2}\right)^{-m}, \quad x \in \mathbb{R},$$

where $m = \frac{5\alpha_4 - 9}{2(\alpha_4 - 3)}$, $\alpha = \sigma_{T_n} \sqrt{\frac{2\alpha_4}{\alpha_4 - 3}} = \sqrt{\frac{2\alpha_4}{\alpha_4 - 3}}$ and $B(a, b)$ is beta function.

Then, cumulative distribution function is equal to

$$F(x) = \frac{1}{2} I_{\alpha^2/(\alpha^2+x^2)}\left(m - \frac{1}{2}, \frac{1}{2}\right),$$

for $x < 0$ and

$$F(x) = 1 - \frac{1}{2} I_{\alpha^2/(\alpha^2+x^2)}\left(m - \frac{1}{2}, \frac{1}{2}\right),$$

for $x > 0$, where $I_x(a, b) = \frac{B_x(a, b)}{B(a, b)}$ and $B_x(a, b)$ is incomplete beta function.

Probability of I type error and power are, respectively equal to

$$\alpha = 2F(-k), \quad 1 - \beta = F(-k - \delta \sqrt{n}) + F(-k + \delta \sqrt{n}).$$

5.3. Johnson SU distribution

In our case, Johnson SU distribution is defined by (Johnson 1949, Johnson et al. 1994)

$$Z = \gamma + \zeta \operatorname{arcsinh}\left(\frac{T_n - \xi}{\psi}\right),$$

where T_n is standardized sample mean (with unbounded range) and $\operatorname{arcsinh}(\cdot)$ is inverse sine hyperbolic function. Transformed variable Z has standard normal distribution. Symbols γ, ζ, ξ and ψ represent parameters of the distribution.

Using method of moments (see Hill, 1976), we get following estimates for parameters γ, ζ, ξ, ψ . We will use notation $\omega = e^{\frac{1}{\zeta^2}}$.

As $\alpha_{3,T_n} = 0$, the required fitted Johnson's curve is symmetric, and

$$\omega = \sqrt{(2\alpha_{4,T_n} - 2)^{\frac{1}{2}} - 1}, \quad \zeta = (\ln \omega)^{-\frac{1}{2}}, \quad \gamma = 0.$$

Parameters ξ and ψ can be found from

$$\xi = \mu_{T_n} = 0, \quad \psi = \sqrt{\frac{2\sigma_{T_n}^2}{\omega^2 - 1}} = \sqrt{\frac{2}{\omega^2 - 1}}.$$

Then, Johnson's transformation is equal to

$$Z = \zeta \operatorname{arcsinh}\left(\frac{T_n}{\psi}\right).$$

Probability of I type error and power are, respectively equal to

$$\begin{aligned} \alpha &= 2\Phi\left(-\zeta \operatorname{arcsinh}\left(\frac{k}{\psi}\right)\right), \\ 1 - \beta &= \Phi\left(\zeta \operatorname{arcsinh}\left(\frac{k + \delta \sqrt{n}}{\psi}\right)\right) - \Phi\left(\zeta \operatorname{arcsinh}\left(\frac{-k + \delta \sqrt{n}}{\psi}\right)\right). \end{aligned}$$

6. Example

Let X represent measurable quality characteristic. Let $a_1 = \$1$, $a_2 = \$0.1$, $a_3 = \$25$, $a'_3 = \$50$, $a_4 = \$100$, $\lambda = 0.05$, $\delta = 2$, $g = 0.0167h$, $D = 1$ (see Duncan, 1956; Montgomery, 2005).

Constrained minimization of expected loss function (1) was done using genetic algorithm *ga* from R package GA (Scrucca, 2013). We selected following values for bounds on type I error probability, power and ATS an expected shift: $\alpha_0 = 0.05$, $p_0 = 0.9$, $ATS_0 = 2h$.

We considered four distributions of quality characteristics: Student distributions with 5 and 10 degrees of freedom, logistic distribution and Laplace distribution with $\eta = 1$. Using results from Section 4, we calculated theoretical distribution of standardized sample mean (or its best approximation) and, afterwards, expressions for type I error probability, power and expected loss function. The same procedure was followed using normal, Pearson VII and Johnson's distributions. We considered samples of sizes 3 to 10, finding, by constrained optimization of expected loss function, the optimal values of parameters k and h .

We provided to genetic algorithm *ga* interval of values for parameters k and h in the following way. Solving equations $\alpha = \alpha(k) = 0.05$ and $p = p(k) = 0.9$ using Brent's root-finding method (Brent, 1973), we got values k_1 and k_2 . By the constraints (1) we got interval of values $[k_1, k_2]$ for parameter k and $[0, 2p(k_1)]$ for parameter h . Results of the minimization of expected loss function are given in Table 1.

By looking at the results in Table 1, we can make following conclusions.

1. Normal approximation of standardized sample mean gives smallest values for k, h , ATS and expected loss and greatest power. One can not be automatically rely on these results, considering the fact that theoretical distribution of standardized sample mean gives the most precise answer.
2. Student distribution with 10 degrees of freedom is very close to normal distribution, so it is expected that results for normal approximation (optimal values of k and h , α , power, ATS and expected loss) should be close to results based on theoretical distribution, which is confirmed. Normal approximation does not fit so well in the case of Student distribution with $\nu = 5$ degrees of freedom, logistic and Laplace distribution.
3. Pearson's VII and Johnson SU approximations give results which are very close to results based on theoretical distribution of standardized sample mean, in all cases.

Distribution	Theoretical distribution					Normal distribution					Pearson VII distribution					Johnson SU distribution										
	k	h	α	$1-\beta$	ATS	EL	k	h	α	$1-\beta$	ATS	EL	k	h	α	$1-\beta$	ATS	EL	k	h	α	$1-\beta$	ATS	EL		
Student- $\nu = 5$	n = 3	2.25	1.00	0.030	0.90	1.11	11.31	2.18	0.98	0.029	0.90	1.09	11.25	2.27	1.00	0.031	0.90	1.11	11.35	2.27	1.00	0.032	0.90	1.11	11.40	
	n = 4	2.77	0.83	0.011	0.90	0.92	10.39	2.71	0.77	0.0067	0.90	0.86	10.13	2.78	0.82	0.012	0.90	0.91	10.46	2.78	0.84	0.013	0.90	0.93	10.50	
	n = 5	3.17	0.73	0.0047	0.91	0.80	10.04	3.05	0.69	0.0023	0.92	0.75	9.84	3.21	0.72	0.0054	0.91	0.80	10.09	3.23	0.72	0.0054	0.90	0.80	10.11	
	n = 6	3.35	0.72	0.0030	0.94	0.76	9.89	3.20	0.70	0.0014	0.96	0.74	9.74	3.40	0.72	0.0035	0.94	0.77	9.93	3.43	0.72	0.0034	0.93	0.77	9.94	
	n = 7	3.52	0.72	0.0020	0.96	0.74	9.83	3.35	0.71	0.0082	0.97	0.73	9.71	3.59	0.72	0.0022	0.96	0.75	9.86	3.59	0.72	0.0022	0.96	0.75	9.86	
	n = 8	3.68	0.72	0.0013	0.97	0.73	9.81	3.49	0.71	0.00049	0.99	0.72	9.73	3.74	0.71	0.0015	0.97	0.74	9.83	3.75	0.71	0.0014	0.97	0.74	9.83	
	n = 9	3.83	0.72	0.00093	0.98	0.73	9.83	3.62	0.71	0.00030	0.99	0.72	9.77	3.90	0.72	0.00097	0.98	0.73	9.84	3.90	0.71	0.00095	0.98	0.73	9.84	
	n = 10	3.98	0.72	0.00063	0.99	0.72	9.87	3.75	0.71	0.00017	0.99	0.72	9.82	4.03	0.72	0.00066	0.99	0.73	9.87	4.03	0.71	0.00064	0.99	0.72	9.87	
	$\nu = 10$	n = 3	2.20	0.98	0.030	0.90	1.09	11.31	2.18	0.98	0.029	0.90	1.09	11.25	2.20	0.99	0.031	0.90	1.10	11.32	2.20	0.99	0.031	0.90	1.10	11.33
		n = 4	2.73	0.78	0.0084	0.90	0.86	10.23	2.71	0.74	0.0067	0.90	0.82	10.13	2.73	0.75	0.0085	0.90	0.83	10.23	2.73	0.77	0.0084	0.90	0.86	10.23
n = 5		3.08	0.71	0.0032	0.92	0.77	9.90	3.05	0.69	0.0023	0.92	0.75	9.84	3.11	0.70	0.0030	0.92	0.76	9.90	3.10	0.70	0.0030	0.92	0.76	9.90	
n = 6		3.26	0.70	0.0018	0.95	0.74	9.78	3.20	0.70	0.0014	0.96	0.74	9.74	3.26	0.70	0.0018	0.95	0.74	9.78	3.26	0.70	0.0018	0.95	0.74	9.78	
n = 7		3.41	0.71	0.0011	0.97	0.73	9.74	3.34	0.71	0.00083	0.97	0.73	9.71	3.41	0.71	0.0011	0.97	0.73	9.74	3.41	0.71	0.0011	0.97	0.73	9.74	
n = 8		3.55	0.71	0.0067	0.98	0.72	9.75	3.49	0.71	0.00049	0.99	0.72	9.73	3.55	0.71	0.00066	0.98	0.72	9.75	3.55	0.71	0.00066	0.98	0.72	9.75	
n = 9		3.69	0.71	0.00041	0.99	0.72	9.78	3.62	0.71	0.00029	0.99	0.72	9.77	3.69	0.71	0.00041	0.99	0.72	9.78	3.69	0.71	0.00041	0.99	0.72	9.78	
n = 10		3.82	0.71	0.00025	0.99	0.72	9.82	3.74	0.72	0.00018	0.99	0.72	9.82	3.82	0.71	0.00025	0.99	0.72	9.83	3.82	0.71	0.00025	0.99	0.72	9.83	
Logistic		n = 3	2.21	1.00	0.031	0.90	1.11	11.34	2.18	1.00	0.029	0.90	1.11	11.25	2.21	1.00	0.031	0.90	1.11	11.32	2.21	1.02	0.031	0.90	1.13	11.33
		n = 4	2.74	0.77	0.0088	0.90	0.85	10.26	2.72	0.74	0.0066	0.90	0.82	10.12	2.73	0.75	0.0089	0.90	0.83	10.26	2.72	0.77	0.0090	0.90	0.85	10.26
	n = 5	3.10	0.71	0.0032	0.92	0.77	9.92	3.04	0.69	0.0023	0.92	0.75	9.84	3.11	0.71	0.0032	0.92	0.77	9.92	3.08	0.70	0.0034	0.92	0.76	9.92	
	n = 6	3.27	0.70	0.0019	0.95	0.74	9.79	3.20	0.70	0.0014	0.96	0.74	9.74	3.27	0.70	0.0019	0.95	0.74	9.79	3.27	0.71	0.0019	0.95	0.74	9.79	
	n = 7	3.42	0.71	0.0011	0.97	0.73	9.75	3.35	0.71	0.00082	0.97	0.73	9.71	3.42	0.71	0.0011	0.97	0.73	9.75	3.42	0.71	0.0011	0.97	0.73	9.75	
	n = 8	3.56	0.71	0.00069	0.98	0.72	9.75	3.49	0.71	0.00049	0.99	0.72	9.73	3.56	0.71	0.00070	0.98	0.72	9.75	3.56	0.71	0.00069	0.98	0.72	9.75	
	n = 9	3.70	0.71	0.00042	0.99	0.72	9.78	3.62	0.71	0.00029	0.99	0.72	9.77	3.70	0.71	0.00043	0.99	0.72	9.78	3.70	0.71	0.00043	0.99	0.72	9.78	
	n = 10	3.83	0.71	0.00026	0.99	0.72	9.83	3.75	0.71	0.00018	0.99	0.72	9.82	3.83	0.71	0.00027	0.99	0.72	9.83	3.83	0.71	0.00027	0.99	0.72	9.83	
	Laplace $\eta = 1$	n = 3	2.24	1.02	0.034	0.90	1.13	11.46	2.18	0.98	0.029	0.90	1.09	11.25	2.24	0.99	0.031	0.90	1.10	11.35	2.24	1.01	0.032	0.90	1.13	11.38
		n = 4	2.76	0.82	0.011	0.90	0.91	10.41	2.71	0.76	0.0066	0.90	0.84	10.13	2.75	0.79	0.011	0.90	0.87	10.37	2.76	0.78	0.011	0.90	0.87	10.38
n = 5		3.18	0.70	0.0041	0.91	0.78	10.00	3.05	0.69	0.0023	0.92	0.75	9.84	3.16	0.71	0.0042	0.91	0.78	10.00	3.17	0.71	0.0042	0.91	0.79	10.00	
n = 6		3.36	0.71	0.0024	0.94	0.75	9.85	3.20	0.70	0.0014	0.96	0.73	9.74	3.34	0.71	0.0026	0.94	0.75	9.86	3.34	0.71	0.0025	0.94	0.76	9.86	
n = 7		3.41	0.71	0.0011	0.97	0.73	9.74	3.34	0.71	0.00083	0.97	0.73	9.71	3.41	0.71	0.0011	0.97	0.73	9.74	3.41	0.71	0.0011	0.97	0.73	9.74	
n = 8		3.55	0.71	0.00067	0.98	0.72	9.75	3.49	0.71	0.00049	0.99	0.72	9.73	3.55	0.71	0.00066	0.98	0.72	9.75	3.55	0.71	0.00066	0.98	0.72	9.75	
n = 9		3.69	0.71	0.00041	0.99	0.72	9.78	3.62	0.71	0.00029	0.99	0.72	9.77	3.69	0.71	0.00041	0.99	0.72	9.78	3.69	0.71	0.00041	0.99	0.72	9.78	
n = 10		3.82	0.71	0.00025	0.99	0.72	9.82	3.74	0.72	0.00018	0.99	0.72	9.82	3.82	0.71	0.00025	0.99	0.72	9.83	3.82	0.71	0.00025	0.99	0.72	9.83	

Table 1: Optimum solution to Example

Our results suggest use of reliable method of economic statistical design of \bar{X} control chart based on Pearson's VII or Johnson's SU approximations in the case when distribution of the data is symmetric but non-normal. Survey of literature (see for instance Li and Lin, 2013; Wu and Govindaraju, 2014; Wu et al., 2014) indicates possibilities of application of our method in statistical quality control.

7. Conclusions

We considered optimum economic statistical design of \bar{X} control chart when quality characteristic has one of the following non-normal symmetric distributions: Student distribution with 5 and 10 degrees of freedom, logistic distribution and Laplace distribution with $\eta = 1$. We calculated theoretical distribution of standardized sample mean (or its best approximation) and approximated it with normal, Pearson type VII and Johnson SU distribution. Constrained minimization of expected loss function was done using genetic algorithm *ga* from R package *GA*. For considered example, we got table of results for optimal values of parameters k and h , values of type I error probability, power, ATS and minimal value of expected loss, for sample sizes from 3 to 10. For all chosen distributions of quality characteristic, Pearson type VII distribution and Johnson SU distribution provide results which are very close to results based on theoretical distribution of standardized sample mean, while normal distribution gives much worse fit.

Although our results of economic statistical design of \bar{X} -bar chart are based only on three types of non-normal symmetric distributions, we would recommend use of Pearson type VII distribution or Johnson SU distribution in general case, when distribution of quality characteristic is non-normal but symmetric. Other cases will be in scope of our future research.

References

- [1] Ahsanullah, M., Golam Kibria B.M. and Shakil, M. (2014). *Normal and Student's t Distributions and Their Applications*, Atlantis Press, Paris.
- [2] Balakrishnan, N. (ed.) (1992). *Handbook of the Logistic Distribution*, Dekker, New York.
- [3] Alloway, J.A. and Raghavachari, M. (1991). Control charts based on Hodges-Lehmann estimator. *Journal of Quality Technology*, 23(4): 336–347.
- [4] Alwan, L.C. (1995). The Problem of Misplaced Control Limits. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(3): 269–278.
- [5] Brent, R.P. (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, New Jersey.
- [6] Duncan, A.J. (1956). The Economic Design of \bar{X} Charts used to Maintain Current Control of a Process. *Journal of the American Statistical Association*, 51(274): 228–242.
- [7] Đorić, D., Mališić, J., Jevremović, V. and Nikolić-Đorić, E. (2007). *Atlas raspodela*. Gradjevinski fakultet, Beograd.
- [8] Gil-Pelaez, J. (1951). Note on the inversion theorem, *Biometrika* 38, pp. 481 - 482.
- [9] Gupta, S.S. and Han, S. (1992). Selection and ranking procedures for logistic populations, In: *Order Statistics and Nonparametrics: Theory and Applications* (Edited by P.K. Sen and I.A. Salama), 377–404. Elsevier, Amsterdam,
- [10] Hill, I.D., Hill, R. and Holder, R.L. (1976). Algorithm AS 99: Fitting Johnson Curves by Moments. *Journal of the Royal Statistical Society - Series C (Applied Statistics)*, 25(2): 180–189.
- [11] Janacek, G. J. and Meikle, S.E. (1997). Control charts based on medians. *The Statistician*, 46 (1): 19–31.
- [12] Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36(1/2): 149–176.
- [13] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions Volume 1*, second edition. Wiley, New York.
- [14] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions Volume 2*, second edition. Wiley, New York.
- [15] Li, D-C. and Lin, L-S. (2013). A new approach to assess product lifetime performance for small data sets. *European Journal of Operational Research*, 230: 290–298.
- [16] Kotz, S., Kozubowski, T.J. and Podgórski, K. (2001). *The Laplace Distribution and Generalizations : a Revisit with Applications to Communications, Economics, Engineering, and Finance*. Springer Science+Business Media, New York.
- [17] Küchler, U. and Tappe, S. (2008). On the shapes of bilateral Gamma densities. *Statistics & Probability Letters* 78(15): 2478–2484.
- [18] Lorenzen, T.J. and Vance L.C. (1986). The Economic Design of Control Charts: A Unified Approach. *Technometrics*, 28(1): 3–10.
- [19] Montgomery, D.C. (2005). *Introduction to Statistical Quality Control*, fifth edition. Wiley, New York.
- [20] Saniga, E.M. (1989). Economic Statistical Control-Chart Design with an Application to \bar{X} and R Charts. *Technometrics*, 31(3): 313–320.
- [21] Scrucca, L. (2013). *GA: A Package for Genetic Algorithms in R*. *Journal of Statistical Software*, 53(4): 1–37.
- [22] Witkovský, V. (2001). On the exact computation of the density and of the quantiles of linear combinations of t and F random variables. *Journal of Statistical Planning and Inference* 94(1): 1–13.

- [23] Witkovský, V. (2004). Matlab algorithm tdist: the distribution of a linear combination of Student's t random variables. *COMPSTAT 2004 Symposium*.
- [24] Witkovsky, V. and Savin, A. (2005). tdist: Distribution of a linear combination of independent Student's t -variables, R package version 0.1.1.
- [25] Woodall, W.H. (1985). The Statistical Design of Quality Control Charts. *Journal of the Royal Statistical Society Series D* 34(2):155–160.
- [26] Wu, H. and Govindaraju, K. (2014). Computer-aided variables sampling inspection plans for compositional proportions and measurement error adjustment. *Computers and Industrial Engineering*, 72: 239–246.
- [27] Wu, J-W., Hong, C-W. and Lee, W-C. (2014). Computational procedure of lifetime performance index of products for the Burr XII distribution with upper record values. *Applied Mathematics and Computation*, 227: 701–716.