X bar control chart for non-normal symmetric distributions

Kristina Veljkovic [1]

Abstract

In statistical quality control, X bar control chart is extensively used to monitor a change in the process mean. In this paper, X bar control chart for non-normal symmetric distributions is proposed. For chosen Student, Laplace, logistic and uniform distributions of quality characteristic, we calculated theoretical distribution of standardized sample mean and fitted Pearson type II or type VII distributions. Width of control limits and power of the X bar control chart were established, giving evidence of the goodness of fit of the corresponding Pearson distribution to the theoretical distribution of standardized sample mean. For implementation of X bar control chart in practice, numerical example of construction of a proposed chart is given.

1 Introduction

The X bar chart is extensively used in practice to monitor a change in the process mean. It is usually assumed that measured quality characteristic has normal or approximately 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). Misplaced 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 the 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 four types of non-normal symmetric distributions of quality characteristic: Student, Laplace, logistic and uniform distributions. These distributions are chosen because of their applications in various disciplines (economics, finance, engineering, hydrology, etc., see for instance Ahsanullah, et al., 2014; Balakrishnan, 1992; Kotz et al., 2001). For each of these distributions, we calculated theoretical distribution of the standardized sample mean (or its best approximation) and approximated it with Pearson type II or type VII distributions. Pearson system of distributions is known to provide approximations to a wide variety of observed distributions (Johnson et al., 1994).

---

[1] Department of Probability and Statistics, Faculty of Mathematics, University of Belgrade, Serbia; kristina@matf.bg.ac.rs

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 the process standard deviation (Montgomery, 2005). It is also assumed that the standard deviation remains stable. Center line of the X bar chart is set at 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 of size n are taken from the process 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 rest of the paper is organized as follows. In Sections 2, 3 and 4, respectively, descriptions of chosen distributions of quality characteristic, distributions of standardized sample mean and Pearson types II and VII distributions are given. Construction of the X bar control chart and its power are examined in Section 5, along with the comparisons of theoretical distribution of sample mean with the corresponding Pearson distribution. In Section 6, implementation of proposed X bar chart is considered. Finally, conclusions are drawn in Section 7.

2   Distribution of quality characteristic

We considered four types of non-normal symmetric distributions of quality characteristic X: Student distribution $t(10)$, standard Laplace $L(1)$ distribution and logistic distribution $LGS(1)$ (see Johnson et al. 1994; Johnson et al. 1995) as representatives of symmetric distributions with heavier tails than normal distribution (Figure 1) and uniform $U(0,1)$ distribution as a representative of symmetric distributions with lighter tails than normal distribution. For simplicity, we have chosen standard forms of all four distributions.
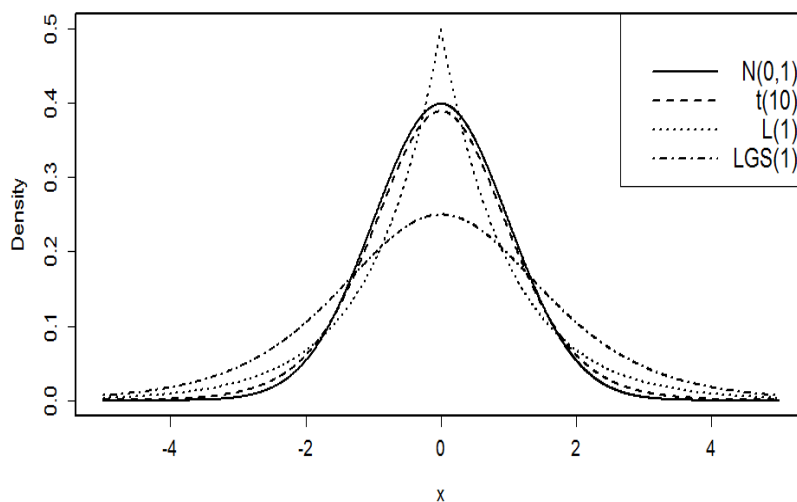

Figure 1: Probability density functions of Student $t(10)$, Laplace $L(1)$, logistic $LGS(1)$ and standard normal $N(0,1)$ distributions

Distribution	f_X	μ	σ^2	α_4
$t(10)$	$\frac{315}{256\sqrt{10}}\left(1+\frac{x^2}{10}\right)^{-5.5}$, $x \in \mathbb{R}$	0	1.25	4
$L(1)$	$\frac{1}{2}e^{-\|x\|}$ $x \in \mathbb{R}$	0	2	6
$LGS(1)$	$\frac{e^{-x}}{(1+e^{-x})^2}$ $x \in \mathbb{R}$	0	$\frac{\pi^2}{3}$	4.2
$U(0,1)$	x, $x \in [0,1]$	0.5	$\frac{1}{12}$	1.8

Table 1: Chosen distributions of quality characteristics

Distributions are given in Table 1 by their probability density function f_X, mean μ, variance $\sigma^2 = Var(X)$ and kurtosis $\alpha_4 = \frac{E(X-E(X))^4}{\sigma^4}$. As all chosen distributions are symmetric around the zero, skewness $\alpha_3 = \frac{E(X-E(X))^3}{\sigma^{\frac{3}{2}}} = 0$.

3 Distribution of standardized sample mean

For chosen distributions of quality characteristic, we will derive the distribution of standardized sample mean $T_n = \frac{\overline{X}-\mu}{\sigma}\sqrt{n}$. As all chosen distributions are symmetric, skewness of standardized sample mean will also be equal to 0.

3.1 Sample from Student's distribution

Witkowský (2001, 2004) proposed a method for numerical evaluation of the distribution function of a linear combination of independent Student variables. The method is based on the inversion formula which leads to the one-dimensional numerical integration.

Let $(X_1, X_2, \ldots, X_n)$ be a sample from Student $t(\nu)$ distribution. Further, let $Y = \sum_{k=1}^{n} X_k$ be sum of these variables and $\phi_{X_k}(t)$ denote the characteristic function of X_k. The characteristic function of Y is

$$\phi_Y(t) = \prod_{k=1}^{n} \phi_{X_k}(t) = \prod_{k=1}^{n} \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)}\left(\nu^{\frac{1}{2}}|t|\right)^{\frac{\nu}{2}} K_{\nu/2}\left(\nu^{\frac{1}{2}}|t|\right),$$

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

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

$$F_Y(y) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin(ty)\phi_Y(t)}{t}\,\mathrm{d}t \qquad (3.1)$$

For any chosen y algorithm $tdist$ in R package $tdist$ (Witkowský and Savin, 2005) evaluates the integral in (3.1) by multiple p-points Gaussian quadrature over the real interval $t \in (0, 10\pi)$. The whole interval is divided in m subintervals and the integration over each subinterval is done with p-points Gaussian quadrature which involves base points

b_{ij}, and weight factors w_{ij}, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, m$. So,

$$F_Y(y) \approx \frac{1}{2} + \frac{1}{\pi} \sum_{j=1}^{m} \sum_{i=1}^{p} \frac{\sin(b_{ij}y)}{b_{ij}} w_{ij} \phi_Y(b_{ij}).$$

Then, cumulative distribution function of standardized sample mean is equal to

$$F_{T_n}(t) = F_Y\left(\frac{\sqrt{5n}}{2} t\right), \ t \in \mathbb{R}.$$

Kurtosis of T_n is equal to $\alpha_{4, T_n} = 3 + \frac{1}{n}$

3.2 Sample from Laplace distribution

Let $(X_1, X_2, \ldots, X_n)$ be a sample from standard Laplace $L(1)$ distribution. Difference of two independent random variables with standard exponential $\varepsilon(1)$ distribution has standard Laplace distribution. Further, standard exponential distribution is gamma distribution, $\Gamma(1,1)$. Sum of n independent variables with $\Gamma(1,1)$ distribution is gamma distribution $\Gamma(n,1)$. In that way, we conclude that sum Y of n independent random variables $X_1, X_2, \ldots, X_n$ with standard Laplace distribution can be written as the difference of two random variables with gamma distribution $\Gamma(n,1)$ which is called bilateral gamma distribution.

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

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

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

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

and $\gamma(n, y)$ is incomplete gamma function.

Then, cumulative distribution function of standardized sample mean is equal to

$$F_{T_n}(t) = F_Y\left(\sqrt{2n}t\right), \ t \in \mathbb{R}.$$

Kurtosis of standardized sample mean is equal $\alpha_{4, T_n} = 3 + \frac{3}{n}$.

3.3 Sample from logistic distribution

Let $(X_1, X_2, \ldots, X_n)$ be a random sample from logistic $LGS(1)$ distribution. Insofar, the best approximation of the distribution of standardized sample mean T_n is given by Gupta and Han (1992). They considered the Edgeworth series expansions up to order n^{-3} for

the distribution of the standardized sample mean. Cumulative distribution function of T_n is given by

$$F_{T_n}(t) \approx \Phi(t) - \varphi(t) \left(\frac{1}{n} \left(\frac{1}{4!} \frac{6}{5} H_3(t) \right) + \frac{1}{n^2} \left(\frac{1}{6!} \frac{48}{7} H_5(t) + \right.\right.$$

$$\left. + \frac{35}{8!} \left(\frac{6}{5} \right)^2 H_7(t) \right) + \frac{1}{n^3} \left(\frac{1}{8!} \frac{432}{5} H_7(t) + \frac{210}{10!} \frac{48}{7} \frac{6}{5} H_9(t) + \right.$$

$$\left.\left. + \frac{5775}{12!} \left(\frac{6}{5} \right)^3 H_{11}(t) \right)\right), \ t \in \mathbb{R},$$

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

Kurtosis of standardized sample mean is $\alpha_{4,T_n} = 3 + \frac{1.2}{n}$.

3.4   Sample from uniform distribution

Let $(X_1, X_2, \ldots, X_n)$ be a random sample from uniform $U(0,1)$ distribution. The sum $Y = \sum_{k=1}^{n} X_k$ has Irwin-Hall distribution (Johnson et al., 1995) with cumulative distribution function

$$F_Y(y) = \frac{1}{2} + \frac{1}{2n!} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \mathrm{sgn}(y-k)(y-k)^n, \ x \in \mathbb{R}.$$

Then, standardized sample mean has cumulative distribution function equal to

$$F_{T_n}(t) = F_Y \left(\left(\frac{t}{\sqrt{12n}} + \frac{1}{2} \right) n \right), \ t \in \mathbb{R}.$$

Kurtosis of standardized sample mean is $\alpha_{4,T_n} = 3 - \frac{1.2}{n}$.

4   Symmetric Pearson distributions

4.1   Pearson type II distribution

Pearson type II distribution can be used for approximation of the distribution of random variable with skewness $\alpha_3 = 0$ and kurtosis $\alpha_4 < 3$ (Johnson et al., 1994). Cumulative distribution function of Pearson type II distribution is equal to

$$F(t) = I_{\frac{t-\lambda}{s}}(a, a), \ 0 < \frac{t-\lambda}{s} < 1,$$

where

$$\lambda = -\sqrt{\frac{2\alpha_4}{3-\alpha_4}}, \quad s = 2\sqrt{\frac{2\alpha_4}{3-\alpha_4}}, \quad a = \frac{5\alpha_4 - 9}{2(3-\alpha_4)} + 1, \tag{4.1}$$

$I_t(a,b) = \frac{B_t(a,b)}{B(a,b)}$, $B(a,b)$ is beta function and $B_t(a,b)$ is incomplete beta function.

In other words, random variable $\frac{T-\lambda}{s}$ has beta distribution $\mathcal{B}(a,a)$.

4.2   Pearson type VII distribution

Pearson type VII distribution can be used for approximation of the distribution of random variable with skewness $\alpha_3 = 0$ and kurtosis $\alpha_4 > 3$ (Johnson et al., 1994). Cumulative distribution function of Pearson type VII distribution is equal to

$$F(t) = \frac{1}{2} I_{a^2/(a^2+t^2)} \left(m - \frac{1}{2}, \frac{1}{2} \right), \ \ t < 0$$

and

$$F(t) = 1 - \frac{1}{2} I_{a^2/(a^2+t^2)} \left(m - \frac{1}{2}, \frac{1}{2} \right), \ \ t > 0,$$

where

$$m = \frac{5\alpha_4 - 9}{2(\alpha_4 - 3)}, \quad a = \sqrt{\frac{2\alpha_4}{\alpha_4 - 3}}. \tag{4.2}$$

5   Design of X bar control chart

For sample sizes $n = 3, 4, \ldots, 10$, we calculated theoretical distribution of the standardized sample mean of considered distributions, using results from Section 3 and then we approximated it with Pearson type II distribution in the case of uniform distribution and with Pearson type VII distribution in the case of Student, Laplace and logistic distributions. Parameters of the fitted Pearson types II and VII distributions are calculated using formulas (4.1) and (4.2). Code for all calculations was written, by the author, in statistical software R and is available as supplementary code on the web site of the Journal. Width of control limits of the X bar control chart is calculated from

$$\alpha = 1 - P\{\mu_0 - k\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu_0 + k\frac{\sigma}{\sqrt{n}} | \mu = \mu_0\} = 2(1 - F_{T_n}(k)), \tag{5.1}$$

where F_{T_n} is cumulative distribution function of standardized sample mean, using Brent's root-finding method (Brent, 1973). Same procedure was followed for both the theoretical distribution of standardized sample mean and corresponding Pearson distribution.

Control limits of the X bar control chart for non-normal symmetric distributions are calculated for specified probability 0.0027 of type I error, in analogy with X bar control chart for normal distribution. When quality characteristics is normally distributed, the probability that sample mean falls outside three standard deviations from the center line is 0.0027, for in-control process. These are so called three-sigma control limits (here sigma refers to the standard deviation of sample mean) and they are frequently used in construction of X bar control chart (Montgomery, 2005).

Calculated widths of control limits, for considered distributions of quality characteristic, sample sizes $n = 3, 4, \ldots, 10$, probability of false alarm $\alpha = 0.0027$, for theoretical distribution of the standardized sample mean and Pearson types II and VII distributions, are given in Table 2.

As it can be seen in the Table 2, the values of the width of the control limits calculated from theoretical distribution and corresponding Pearson distribution are very close, i.e. corresponding Pearson distribution fits very well to the theoretical distribution of the

Sample	Width of control limits							
	Student $t(10)$		*Laplace* $L(1)$		*Logistic* $LGS(1)$		*Uniform* $U(0,1)$	
size	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson
$n=3$	3.21966	3.22227	3.54221	3.53915	3.25580	3.26074	2.59834	2.65308
$n=4$	3.16998	3.17156	3.43224	3.43628	3.20035	3.20234	2.72926	2.74902
$n=5$	3.13867	3.13966	3.36034	3.36606	3.16405	3.16527	2.79650	2.80355
$n=6$	3.11712	3.11775	3.30939	3.31520	3.13877	3.13966	2.83511	2.83866
$n=7$	3.10136	3.10178	3.27130	3.27668	3.12021	3.12091	2.86060	2.86314
$n=8$	3.08934	3.08962	3.24168	3.24652	3.10602	3.10660	2.87932	2.88118
$n=9$	3.07987	3.08005	3.21796	3.22227	3.09482	3.09531	2.89366	2.89502
$n=10$	3.07221	3.07233	3.19852	3.20234	3.08577	3.08619	2.90489	2.90597

Table 2: Width of control limits of X bar control chart

standardized sample mean. On the other hand, normal approximation would give value of $k = 2.99998$, for all n and all distributions of quality characteristics.

Now, we are interested to see what is the power of X bar control charts for detecting shifts $\delta = 0.5, 1.0, \ldots, 3.0$, for calculated width of control limits. Power of X bar control chart for detecting shifts from mean μ_0 to $\mu_1 = \mu_0 \pm \delta\sigma$ can be calculated from

$$
\begin{aligned}
1 - \beta &= 1 - P\{\mu_0 - k\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu_0 + k\frac{\sigma}{\sqrt{n}}|\mu = \mu_1\} = \\
&= F_{T_n}(-k - \delta\sqrt{n}) + F_{T_n}(-k + \delta\sqrt{n}).
\end{aligned}
$$

We should note that power of proposed X bar control chart for detecting shift $\delta = 0$ is 0.0027 for all considered distributions and sample sizes, i.e. it maintains probability of type I error.

Mainly, we want to investigate what is the minimum shift that X bar control chart can detect with a power of at least 90%.

Calculated power of X bar control chart, for considered distributions of quality characteristic, sample sizes $n = 3, 4, \ldots, 10$, shifts $\delta = 0.5, 1.0, \ldots, 3.0$ for both theoretical distribution of standardized sample mean and corresponding Pearson distribution, are given in Table 3.

¿From the Table 3, we see that X bar control chart can detect shifts of $\delta = 1.5$ with power of at least 90% for sample sizes of $n = 9$ and greater for all considered distributions. In order for the X bar chart to detect shifts of $\delta = 2.0$ with power of 90% and greater, it is necessary to take samples of size at least $n = 4$ for Student, Laplace and logistic distributions and sample sizes of $n = 5$ and greater for uniform distribution of quality characteristic. Also, we can once more notice that the corresponding Pearson distribution approximates the distribution of standardized sample mean rather well. In general, it can be concluded that X bar control chart can detect shifts of at least $\delta = 1.5$ with power of 90% and greater for non-normal symmetric distribution of quality characteristic.

6   Implementation of proposed X bar control chart

Now we are interested to see how proposed X bar control chart can be implemented in practice, in case when the distribution function of the quality characteristic is non-normal,

Distribution	Power										
	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson
$t(10)$	$\delta = 0.5$		$\delta = 1.0$		$\delta = 1.5$		$\delta = 2.0$		$\delta = 2.5$		$\delta = 3.0$	
$n = 3$	0.0110	0.0110	0.0665	0.06620	0.2609	0.2597	0.5998	0.5989	0.8715	0.8711	0.9750	0.9748
$n = 4$	0.0162	0.0162	0.1175	0.1171	0.4307	0.4300	0.8016	0.8014	0.9661	0.9659	0.9967	0.9967
$n = 5$	0.0224	0.0224	0.1795	0.1791	0.5871	0.5868	0.9108	0.9107	0.9919	0.9918	0.9996	0.9996
$n = 6$	0.0296	0.0296	0.2488	0.2484	0.7145	0.7144	0.9625	0.9625	0.9982	0.9982	1	1
$n = 7$	0.0377	0.0377	0.3218	0.3216	0.8100	0.8100	0.9850	0.9850	0.9996	0.9996	1	1
$n = 8$	0.0468	0.0468	0.3957	0.3955	0.8775	0.8775	0.9942	0.9942	0.9999	0.9999	1	1
$n = 9$	0.0566	0.0566	0.4678	0.4677	0.9231	0.9231	0.9979	0.9978	1	1	1	1
$n = 10$	0.0674	0.0673	0.5363	0.5363	0.9528	0.9528	0.9992	0.9992	1	1	1	1
$L(1)$												
$n = 3$	0.0077	0.0071	0.0370	0.0355	0.1541	0.1588	0.4642	0.4674	0.8061	0.8014	0.9514	0.9532
$n = 4$	0.0111	0.0104	0.0718	0.0708	0.3179	0.3207	0.7317	0.7258	0.9433	0.9438	0.9916	0.9921
$n = 5$	0.0156	0.0148	0.1215	0.1212	0.4973	0.4949	0.8761	0.8740	0.9841	0.9846	0.9986	0.9986
$n = 6$	0.0209	0.0201	0.1842	0.1842	0.6511	0.6469	0.9452	0.9451	0.9957	0.9958	0.9998	0.9997
$n = 7$	0.0273	0.0265	0.2561	0.2559	0.7670	0.7638	0.9765	0.9767	0.9989	0.9988	1	0.9999
$n = 8$	0.0347	0.0339	0.3327	0.3321	0.8487	0.8469	0.9901	0.9903	0.9997	0.9997	1	1
$n = 9$	0.0430	0.0423	0.4101	0.4089	0.9039	0.90305	0.9960	0.9960	0.9999	0.9999	1	1
$n = 10$	0.0523	0.0517	0.4850	0.4835	0.9400	0.9397	0.9984	0.9983	1	1	1	1
$LGS(1)$												
$n = 3$	0.0106	0.0103	0.0619	0.0613	0.2469	0.2460	0.5864	0.5840	0.8653	0.8638	0.9726	0.9726
$n = 4$	0.0155	0.0153	0.1109	0.1106	0.4178	0.4172	0.7945	0.7935	0.9637	0.9637	0.9962	0.9963
$n = 5$	0.0215	0.0213	0.1718	0.1718	0.5775	0.5769	0.9072	0.9070	0.9911	0.9911	0.9995	0.9995
$n = 6$	0.0285	0.0283	0.2409	0.2408	0.7079	0.7074	0.9607	0.9607	0.9979	0.9979	0.9999	0.9999
$n = 7$	0.0364	0.0363	0.3143	0.3142	0.8056	0.8053	0.9841	0.9842	0.9995	0.9995	1	1
$n = 8$	0.0452	0.0451	0.3887	0.3886	0.8746	0.8744	0.9938	0.9938	0.9999	0.9999	1	1
$n = 9$	0.0550	0.0549	0.4616	0.4614	0.9211	0.9211	0.9977	0.9977	1	1	1	1
$n = 10$	0.0655	0.0655	0.5309	0.5308	0.9515	0.9515	0.9991	0.9991	1	1	1	1
$U(0,1)$												
$n = 3$	0.0424	0.0357	0.2022	0.1890	0.4999	0.4794	0.7976	0.7803	0.9575	0.9531	0.9986	0.9976
$n = 4$	0.0419	0.0397	0.2403	0.2348	0.6030	0.5951	0.8937	0.8898	0.9906	0.9903	1	1
$n = 5$	0.0470	0.0461	0.2929	0.2909	0.7062	0.7034	0.9527	0.9521	0.9986	0.9985	1	1
$n = 6$	0.0544	0.0538	0.3533	0.3523	0.7946	0.7934	0.9816	0.9816	0.9999	0.9998	1	1
$n = 7$	0.0630	0.0626	0.4167	0.4159	0.8629	0.8622	0.9936	0.9936	1	1	1	1
$n = 8$	0.0726	0.0723	0.4801	0.4794	0.9120	0.9117	0.9980	0.9980	1	1	1	1
$n = 9$	0.0830	0.0828	0.5416	0.5411	0.9455	0.9453	0.9994	0.9994	1	1	1	1
$n = 10$	0.0942	0.0941	0.6001	0.5996	0.9673	0.9672	0.9999	0.9998	1	1	1	1

Table 3: Power of X bar control chart

symmetric but unknown. For fitting Pearson type II or type VII distributions to data, we need an estimate of kurtosis based on sample of means.

6.1 Measures of sample kurtosis

We have three measures of sample kurtosis

$$g_2^* = \frac{m_4}{m_2^2}, \quad G_2^* = \frac{N-1}{(N-2)(N-3)}\left((N+1)g_2 + 6\right) + 3, \quad b_2^* = \frac{m_4}{s^4},$$

where m_k are sample central moments.

Joanes and Gill (1998) investigated three measures $g_2 = g_2^* - 3$, $G_2 = G_2^* - 3$ and $b_2 = b_2^* - 3$ of sample excess kurtosis. They showed that, generating 100000 samples of different sizes from Student t_5 distribution, g_2 generally has the smallest mean-squared error. We followed the same procedure for measures g_2^*, G_2^* and b_2^* and generated 100000 samples of different sizes from distributions of standardized sample mean of Student $t(10)$, Laplace $L(1)$, logistic $LGS(1)$ and uniform $U(0,1)$ distributions. We confirm Joanes and Gill's findings. So, we will use, for calculation of the parameters of Pearson types II and VII distributions, measure g_2^* as an estimate of sample kurtosis.

6.2 Empirical power of X bar control chart

In this section, we will calculate the empirical power of proposed X bar control chart in order to investigate its performance in practice. We will take, by Monte Carlo simulations, $m = 25, 50, 100$ samples of sizes 3 to 10 from Student t(10), Laplace $L(1)$, logistic $LGS(1)$ and uniform $U(0,1)$ distributions. Sample means, as well as estimates of mean and standard deviation, are calculated. Further, we estimated kurtosis of the distribution of sample mean with g_2^*. Then, corresponding Pearson type II or type VII distribution is fitted to m sample means and control limits and power of the X bar control chart are calculated. This procedure is repeated 100000 times. The average power of the X bar control chart, for considered distributions, is presented in Table 4 (rounded to four decimal places). It is expected that sample size and number of groups will affect sample estimates, i.e. values of parameters of fitted Pearson distribution and therefore power of proposed X bar control chart.

We compared the values of empirical power for a number of groups $m = 25, 50, 100$ with theoretical power from Table 3, giving accent on the values of theoretical power of 90% and greater. We made the following conclusions for shift sizes of 1.5 and greater. Zero difference is present at sample sizes of at least 7 and $\delta = 3$. Absolute difference between theoretical and empirical power gets smaller as a number of groups and shift sizes rise. In most of the cases, the difference exists on third to the fourth decimal place. In other words, proposed X bar control chart has quite satisfactory performance. General advice for its use in practice would be to choose preferably more than 25 groups of sample size of 9 and greater, in order to detect shift $\delta = 1.5$ with the power of at least 90%.

Distribution	Power																	
	$m = 25$						$m = 50$						$m = 100$					
$t(10)$	$\delta = 0.5$	$\delta = 1.0$	$\delta = 1.5$	$\delta = 2.0$	$\delta = 2.5$	$\delta = 3.0$	$\delta = 0.5$	$\delta = 1.0$	$\delta = 1.5$	$\delta = 2.0$	$\delta = 2.5$	$\delta = 3.0$	$\delta = 0.5$	$\delta = 1.0$	$\delta = 1.5$	$\delta = 2.0$	$\delta = 2.5$	$\delta = 3.0$
$n = 3$	0.0522	0.1857	0.4231	0.7150	0.9163	0.9842	0.0302	0.1369	0.3643	0.6701	0.8986	0.9808	0.0201	0.1073	0.3252	0.6423	0.8873	0.9786
$n = 4$	0.0688	0.2541	0.5670	0.8627	0.9776	0.9972	0.0424	0.2014	0.5161	0.8378	0.9734	0.9968	0.0292	0.1672	0.4822	0.8219	0.9705	0.9967
$n = 5$	0.0852	0.3225	0.6914	0.9394	0.9937	0.9993	0.0551	0.2688	0.6495	0.9281	0.9928	0.9993	0.0391	0.2328	0.6233	0.9206	0.9924	0.9993
$n = 6$	0.1019	0.3909	0.7909	0.9742	0.9980	0.9998	0.0682	0.3376	0.7588	0.9699	0.9979	0.9998	0.0499	0.3014	0.7385	0.9667	0.9978	0.9998
$n = 7$	0.1180	0.4565	0.8629	0.9888	0.9993	0.9999	0.0821	0.4060	0.8405	0.9872	0.9992	0.9999	0.0615	0.3717	0.8264	0.9863	0.9993	0.9999
$n = 8$	0.1341	0.5199	0.9128	0.9949	0.9997	1.0000	0.0963	0.4728	0.8984	0.9944	0.9997	1.0000	0.0736	0.4405	0.8886	0.9942	0.9997	1.0000
$n = 9$	0.1517	0.5820	0.9460	0.9976	0.9999	1.0000	0.1105	0.5360	0.9366	0.9974	0.9999	1.0000	0.0865	0.5075	0.9307	0.9975	0.9999	1.0000
$n = 10$	0.1679	0.6380	0.9664	0.9988	0.9999	1.0000	0.1253	0.5958	0.9609	0.9987	0.9999	1.0000	0.0996	0.5702	0.9575	0.9988	0.9999	1.0000
$L(1)$																		
$n = 3$	0.0364	0.1420	0.3542	0.6481	0.8828	0.9751	0.0188	0.0930	0.2814	0.5841	0.8531	0.9677	0.0119	0.0655	0.2325	0.5401	0.8326	0.9622
$n = 4$	0.0528	0.2117	0.5102	0.8263	0.9681	0.9956	0.0292	0.1536	0.4427	0.7886	0.9596	0.9944	0.0187	0.1172	0.3953	0.7628	0.9533	0.9936
$n = 5$	0.0691	0.2828	0.6481	0.9227	0.9911	0.9990	0.0408	0.2205	0.5913	0.9040	0.9889	0.9988	0.0271	0.1795	0.5532	0.8913	0.9873	0.9986
$n = 6$	0.0861	0.3539	0.7587	0.9667	0.9972	0.9997	0.0535	0.2912	0.7158	0.9589	0.9966	0.9996	0.0366	0.2483	0.6870	0.9532	0.9962	0.9996
$n = 7$	0.1030	0.4235	0.8412	0.9856	0.9990	0.9999	0.0665	0.3617	0.8101	0.9824	0.9988	0.9999	0.0471	0.3203	0.7903	0.9802	0.9988	0.9999
$n = 8$	0.1197	0.4902	0.8986	0.9935	0.9996	0.9999	0.0804	0.4316	0.8776	0.9922	0.9995	0.9999	0.0585	0.3927	0.8645	0.9915	0.9995	1.0000
$n = 9$	0.1364	0.5542	0.9367	0.9969	0.9998	1.0000	0.0947	0.4993	0.9232	0.9964	0.9998	1.0000	0.0708	0.4638	0.9148	0.9962	0.9998	1.0000
$n = 10$	0.1536	0.6136	0.9605	0.9984	0.9999	1.0000	0.1100	0.5643	0.9527	0.9982	0.9999	1.0000	0.0833	0.5302	0.9470	0.9982	0.9999	1.0000
$LGS(1)$																		
$n = 3$	0.0495	0.1787	0.4130	0.7058	0.9120	0.9831	0.0284	0.1301	0.3526	0.6590	0.8932	0.9793	0.0185	0.0998	0.3105	0.6281	0.8803	0.9766
$n = 4$	0.0664	0.2483	0.5599	0.8584	0.9766	0.9970	0.0402	0.1943	0.5060	0.8314	0.9717	0.9966	0.0275	0.1600	0.4711	0.8151	0.9686	0.9964
$n = 5$	0.0831	0.3177	0.6864	0.9376	0.9935	0.9993	0.0531	0.2625	0.6425	0.9253	0.9924	0.9992	0.0373	0.2255	0.6147	0.9172	0.9918	0.9993
$n = 6$	0.0996	0.3862	0.7870	0.9734	0.9979	0.9998	0.0662	0.3315	0.7533	0.9685	0.9977	0.9998	0.0481	0.2953	0.7332	0.9654	0.9977	0.9998
$n = 7$	0.1160	0.4524	0.8605	0.9884	0.9992	0.9999	0.0801	0.4005	0.8370	0.9867	0.9992	0.9999	0.0594	0.3650	0.8222	0.9857	0.9993	0.9999
$n = 8$	0.1331	0.5175	0.9117	0.9948	0.9997	1.0000	0.0940	0.4675	0.8959	0.9942	0.9997	1.0000	0.0716	0.4348	0.8859	0.9940	0.9997	1.0000
$n = 9$	0.1500	0.5792	0.9450	0.9975	0.9999	1.0000	0.1085	0.5314	0.9349	0.9973	0.9999	1.0000	0.0847	0.5028	0.9290	0.9973	0.9999	1.0000
$n = 10$	0.1669	0.6359	0.9658	0.9987	0.9999	1.0000	0.1239	0.5930	0.9602	0.9987	0.9999	1.0000	0.0976	0.5662	0.9565	0.9988	0.9999	1.0000
$U(0,1)$																		
$n = 3$	0.0788	0.2527	0.5190	0.7998	0.9553	0.9938	0.0573	0.2227	0.4978	0.7874	0.9543	0.9949	0.0463	0.2061	0.4876	0.7826	0.9538	0.9959
$n = 4$	0.0918	0.3107	0.6377	0.9049	0.9879	0.9988	0.0669	0.2772	0.6157	0.8975	0.9885	0.9991	0.0536	0.2578	0.6054	0.8936	0.9893	0.9994
$n = 5$	0.1068	0.3731	0.7436	0.9584	0.9965	0.9997	0.0784	0.3372	0.7222	0.9557	0.9969	0.9998	0.0627	0.3157	0.7122	0.9542	0.9975	0.9999
$n = 6$	0.1219	0.4356	0.8276	0.9823	0.9989	0.9999	0.0914	0.4002	0.8103	0.9819	0.9991	0.9999	0.0733	0.3779	0.8012	0.9818	0.9993	1.0000
$n = 7$	0.1377	0.4976	0.8886	0.9924	0.9996	1.0000	0.1044	0.4623	0.8759	0.9925	0.9997	1.0000	0.0844	0.4400	0.8685	0.9928	0.9998	1.0000
$n = 8$	0.1525	0.5554	0.9290	0.9965	0.9998	1.0000	0.1181	0.5226	0.9215	0.9967	0.9999	1.0000	0.0968	0.5023	0.9167	0.9971	0.9999	1.0000
$n = 9$	0.1690	0.6132	0.9562	0.9983	0.9999	1.0000	0.1318	0.5801	0.9514	0.9985	0.9999	1.0000	0.1094	0.5614	0.9487	0.9987	1.0000	1.0000
$n = 10$	0.1849	0.6659	0.9729	0.9991	1.0000	1.0000	0.1469	0.6359	0.9706	0.9993	1.0000	1.0000	0.1226	0.6179	0.9691	0.9994	1.0000	1.0000

Table 4: Empirical power of X bar control chart

6.3 Example

Montgomery (2005) gave data set on thickness of a printed circuit board (in inches), for 25 samples of three boards each.
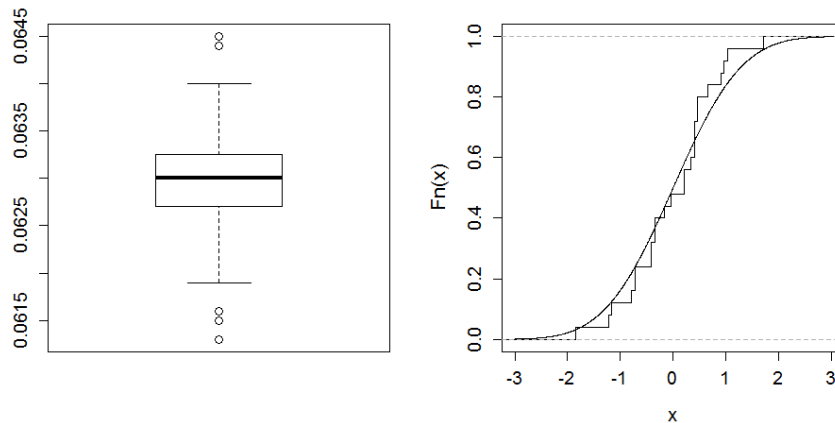

Figure 2: Boxplot of the thickness data (left graph) and empirical cumulative distribution function of standardized sample means with fitted Pearson type II distribution (right graph)

As we can see on boxplot (Figure 2, left graph), sample distribution seems symmetric. We tested symmetry of data distribution using Mira test (Mira, 1999), the Cabilio-Masaro test (Cabilio and Masaro, 1996) and Miao-Gel-Gastwirth (MGG) test (Miao et al., 2006). Based on results of all three tests, we can conclude that data distribution is symmetric (Mira test: Test Statistic = 0.9029, p-value = 0.3666; Cabilio-Masaro test: Test Statistic = 0.8846, p-value = 0.3764; MGG test: Test Statistic = 1.0162, p-value = 0.3095). R function *symmetry.test* for these tests can be found in R package *lawstat* (Gastwirth et al., 2015).

Now we will test the normality of the sample distribution using Shapiro-Wilk, Anderson-Darling and Lilliefors normality tests (Razali and Wah, 2011). Based on results of all three tests, we conclude that data distribution is not normal (Shapiro-Wilk test: W = 0.9589, p-value = 0.01584; Anderson-Darling test: A = 1.4759, p-value = 0.00076; Lilliefors test D = 0.1467, p-value = 0.00039). We used R function *shapiro.test* (package *stats*) for Shapiro-Wilk test and *ad.test*, *lillie.test* from R package *nortest* (Gross and Ligges, 2015) for Anderson-Darling and Lilliefors normality tests, respectively.

For each of 25 samples, we calculated sample mean. Mean of all sample means is equal to $\overline{\overline{X}} = 0.06295$ and this is the estimate of unknown process mean and center line of X bar control chart. Further, we estimated process standard deviation with mean range, $\hat{\sigma} = \overline{R} = 0.00092$. Now, we can calculate standardized sample means and kurtosis of standardized sample means. We got $\hat{\alpha_4} = g_2^* = 2.83154$ (measures of sample excess kurtosis can be found in R package $e1071$ (Meyer et al., 2014)). So, as the distribution of standardized sample means is symmetric with kurtosis smaller than 3, we will approximate its distribution with Pearson type II distribution. We calculated parameters of distribution using equation (4.1). Empirical distribution function along with fitted Pearson

type II distribution of standardized sample means is given on Figure 2, right graph.

For probability of false alarm $\alpha = 0.0027$, we get, using equation (5.1), that width of control limits is equal to $k = 2.83665$. Now we may calculate lower and upper control limits of X bar control chart, $LCL = \overline{\overline{X}} - k\frac{\overline{R}}{\sqrt{n}} = 0.06143$, $UCL = \overline{\overline{X}} + k\frac{\overline{R}}{\sqrt{n}} = 0.06448$ and construct X bar chart (Figure 3). As we can see on Figure 3, all sample means are within the control limits and we can conclude that process is in-control and keep the estimates of unknown process mean, standard deviation, as well the width of control limits.
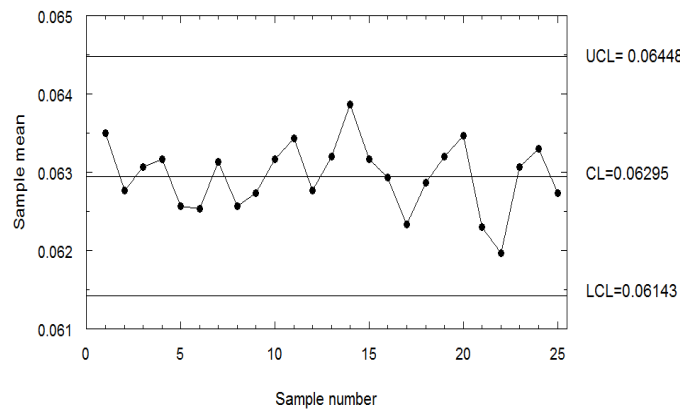

Figure 3: X bar control chart for the thickness data

7   Conclusions

We considered design of the X bar control chart when quality characteristic has one of the following non-normal symmetric distributions: Student distribution with 10 degrees of freedom, standard Laplace, standard logistic and standard uniform distributions. We calculated theoretical distribution of the standardized sample mean (or its best approximation) and approximated it with Pearson type II or type VII distributions. Then we calculated width of control limits of the X bar chart, which gave evidence of the goodness of fit of the corresponding Pearson distribution to the theoretical distribution of the standardized sample mean. Further, we examined the power of X bar control chart in detecting the shifts. Results suggest that the X bar chart can detect shifts of at least $\delta = 1.5$ with power of 90% and greater. Then we undertook Monte Carlo study in order to calculate empirical power of proposed X bar control chart, confirming its quite satisfactory performance. Finally, we constructed X bar chart for a given data set, when data distribution is non-normal and symmetric, but unknown.

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] Alloway, J.A. and Raghavachari, M. (1991): Control charts based on Hodges-Lehmann estimator. *Journal of Quality Technology*, **23**, 336-347.

[3] Alwan, L.C. (1995): The Problem of Misplaced Control Limits, *Journal of the Royal Statistical Society, Series C*, **44**, 269-278.

[4] Balakrishnan, N. (1992): *Handbook of the Logistic Distribution*. Marcel Dekker, New York.

[5] Brent, R.P. (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, New Jersey.

[6] Cabilio, P. and Masaro, J. (1996): A simple test of symmetry about an unknown median.*The Canadian Journal of Statistics*, **24**, 349-361.

[7] Gastwirth, J.L., Gel, Y.R., Wallace Hui, W.L., Miao, W. and Noguchi, K. (2015): *lawstat: Tools for Biostatistics, Public Policy, and Law*. R package version 2.5.

[8] Gil-Pelaez, J. (1951): Note on the inversion theorem. *Biometrika*, **38**, 481-482.

[9] Gross, J. and Ligges, U. (2015): *nortest: Tests for Normality*. R package version 1.0-3.

[10] 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). Elsevier, Amsterdam, 377-404.

[11] Janacek, G.J. and Meikle, S.E. (1997): Control charts based on medians. *The Statistician*, **46**, 19-31.

[12] Joanes, D.N. and Gill, C.A. (1998): Comparing measures of sample skewness and kurtosis. *The Statistician*, **47**, 183-189.

[13] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994): *Continuous Univariate Distributions Volume 1*. Wiley, New York.

[14] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995): *Continuous Univariate Distributions Volume 2*. Wiley, New York.

[15] 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, New York.

[16] Küchler, U. and Tappe, S. (2008): On the shapes of bilateral Gamma densities. *Statistics & Probability Letters*, **78**, 2478-2484.

[17] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. and Leisch, F. (2014): *e1071: Misc Functions of the Department of Statistics, TU Wien*. R package version 1.6-4.

[18] Miao, W., Gel, Y.R. and Gastwirth, J.L. (2006): A New Test of Symmetry about an Unknown Median, In: *Random Walk, Sequential Analysis and Related Topics - A Festschrift in Honor of Yuan-Shih Chow* (Edited by A. Hsiung, C.-H. Zhang and Z. Ying). World Scientific Publisher, Singapore.

[19] Mira, A. (1999): Distribution-free test for symmetry based on Bonferroni's measure. *Journal of Applied Statistics*, **26**, 959-972.

[20] Montgomery, D.C. (2005): *Introduction to Statistical Quality Control*. Wiley, New York.

[21] Razali, N. and Wah, Y.B. (2011): Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, **2**, 21-33.

[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**, 1-13.

[23] Witkovský, V. (2004): Matlab algorithm tdist: the distribution of a linear combination of Student's t random variables. *COMPSTAT 2004 Symposium*, Prague.

[24] Witkovský, V. and Savin, A. (2005): *tdist: Distribution of a linear combination of independent Student's t-variables*. R package version 0.1.1.