Serdica
Mathematical Journal
Сердика
Математическо списание

Serdica Math. J. **42** (2016), 187–202

Serdica
Mathematical Journal

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

R CONTROL CHART FOR POSITIVELY SKEWED DISTRIBUTIONS

Kristina Veljkovic

Communicated by N. M. Yanev

Abstract. In this paper, R control chart for positively skewed distributions is designed. For chosen exponential, gamma and Weibull distributions of quality characteristic, we calculated theoretical distribution of sample range and fitted corresponding Pearson distribution. Upper control limits and power of the R control chart were established, giving evidence of the goodness of fit of Pearson distributions to the theoretical distribution of sample range. For implementation of R control chart in practice, measures of sample skewness and kurtosis were compared and numerical example of construction of a proposed chart was given.

Introduction. Mean range is often employed as an estimate of dispersion in place either of the standard deviation (root mean square) or the mean deviation, when dealing with samples containing only a small number of observations. Tippett [24] has calculated the mean, the standard deviation and the

2010 *Mathematics Subject Classification:* 62E17, 62F10, 62G30, 62P30.
Key words: R control chart, Pearson distribution, positively skewed distribution, sample kurtosis, sample skewness.

moment quotients for the range of the normal distribution up to $n = 1000$. The probability integral of the range was first tabulated by Pearson and Hartley [22], for n up to 20, to four decimal places. Harter [12] computed percentage points of relative (standardized) range to 6 decimal places for normal samples of size $n = 2(1)20(2)40(10)100$, as well as the moments to 8 or more significant figures for normal samples up to $n = 100$.

In statistical quality control, R control chart is often used to monitor process variability when sample size $n \leq 10$ [19]. The upper control limit UCL, center line CL and lower control limit LCL of the R control chart are, respectively, equal to

$$\begin{aligned} UCL &= \overline{R} + 3d_3\frac{\overline{R}}{d_2}, \\ CL &= \overline{R}, \\ LCL &= \overline{R} - 3d_3\frac{\overline{R}}{d_2}, \end{aligned}$$

where $\overline{R}$ is the mean range, d_2 and d_3 are, respectively, the expected value and standard deviation of the relative range. Tables of d_2 and d_3 versus n are reproduced in textbooks on statistical quality control, to facilitate the construction of R control chart. Calculation of table values is based on the assumption that quality characteristic is normal or approximately normal. Mahoney [18] investigated the effects of the coefficient d_2 on the R chart. The d_2 values for the uniform, triangular, exponential and Erlang distributions were derived and compared with those for the normal distribution. The same procedure was done for the coefficient d_3 in the paper by Kao and Ho [17]. In both studies, they demonstrated that the use of d_2 and d_3 values computed based on an normality assumption seriously affects the performance of range charts. Several other researchers also dealt with the problem of non-normality in control charting procedure. Ferrel [10] proposed the geometric mid-range and geometric range charts for a lognormal distribution instead of the Shewhart X bar and R charts. Nelson [20] derived control limits of median, range, scale and location charts for the Weibull distribution. Some researchers proposed heuristic methods for setting the limits of control charts: weighted variance (WV) method [7, 3] weighted standard deviation (WSD) method [6], skewness correction (SC) method [5], adjusted weighted variance (AWSD) method [25].

Occurrence of non-normal data in industry is quite common [1, 13]. Violation of normality assumption results in incorrect control limits of control charts [2]. 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.

Positively skewed distributions are frequently used for fitting manufacturing process data. In construction of R control chart, we considered exponential, gamma and Weibull distributions of quality characteristic. The field of applications of these distributions is vast and encompasses nearly all scientific disciplines (engineering, hydrology, economics, medicine, etc., see for instance [21, 23]). We derived the theoretical distribution of sample range and approximated it with corresponding Pearson distribution. A Pearson system of distributions is known to provide approximations to a wide variety of observed distributions [15].

It is presumed that a process begins in in-control state with standard deviation σ_0 and that single assignable cause results in a change of process standard deviation σ_0 to $\sigma_1 = \nu\sigma_0$, where $\nu > 1$. It is assumed that the process mean remains stable. Center line of the R control chart is set at σ_0 and an upper control limit (lower control limit is zero) is calculated for specified probability of false alarms. Samples of size n are taken from the process and the sample range is plotted on the R control chart. If a sample range exceeds upper control limit, it is assumed that some change in the process standard deviation has occurred and a search for the assignable cause is initiated.

The rest of the paper is organized as follows. Construction of the proposed R control chart and its power are examined in Section 2, along with the comparisons of theoretical distribution of sample range with the corresponding Pearson distribution. In Section 3, implementation of proposed R control chart is considered, first by choosing appropriate measures of sample skewness and kurtosis necessary for fitting Pearson distribution to data of sample ranges, followed by numerical illustration, using a data set given by Dou and Sa [8]. Finally, conclusions are drawn in Section 4. Chosen distributions of quality characteristic, distribution of sample range and the Pearson system of distributions are described in Appendix.

2. R control chart. We will consider three positively skewed distributions of quality characteristic X: standard exponential, Gamma $\Gamma(2,1)$ and Weibull $W(2,1)$ distributions [9, 15, 16]. Parameters of these distributions were chosen to provide various degrees of skewness and kurtosis of parent distribution and, also for the ease of computation of theoretical distribution of sample range.

Coefficients of skewness and kurtosis are defined as $\alpha_3 = \dfrac{E\left(X - E(X)\right)^3}{\sigma^{\frac{3}{2}}}$ and $\alpha_4 = \dfrac{E\left(X - E(X)\right)^4}{\sigma^4}$, respectively. We have following values of α_3 and α_4

for chosen distributions.

1. Exponential distribution: $\alpha_3 = 2,\ \alpha_4 = 9,$
2. Gamma $\Gamma(2,1)$ distribution:$\alpha_3 = \sqrt{2},\ \alpha_4 = 6$
3. Weibull $W(2,1)$ distribution: $\alpha_3 = \dfrac{2\sqrt{\pi}(\pi-3)}{(4-\pi)^{1.5}} \approx 0.631,\ \alpha_4 = \dfrac{32-3\pi^2}{(4-\pi)^2} \approx$ 3.245

We will now calculate the theoretical distribution of a sample range of considered positively skewed distributions, for sample sizes n from 3 to 10, using expression (2). After that, we will calculate first four moments of the distribution of sample range, which are necessary for calculation of parameters of the corresponding Pearson distribution. In the case of standard exponential and Gamma $G(2,1)$ distributions, distribution of sample range is approximated with Pearson type VI distribution. For the distribution of a sample range of Weibull $W(1)$, Pearson type I distribution is used for approximation for sample sizes from 3 to 7 and Pearson type VI distribution for sample sizes from 8 to 10. Parameters of corresponding Pearson type I and type VI distributions are calculated using formulas (4),(9) and (11), given in the Appendix. For fitting Pearson distributions, we used functions in R package *PearsonDS* and the rest of the code for all further calculations was written, by the author, in R.

For probability of false alarms $\alpha = 0.0027$, upper control limit UCL of R control chart for both theoretical distribution of sample range and its Pearson distribution can be calculated from

$$\alpha = P\{R \geq UCL|\sigma = \sigma_0\} = 1 - F(UCL), \tag{1}$$

where R is the sample range and F its cumulative distribution function (theoretical distribution function or corresponding Pearson distribution function). Then, upper control limit of R chart is found using Brent's root-finding method [4].

Calculated upper control limits, for sample sizes from 3 to 10, probability of false alarms $\alpha = 0.0027$, theoretical distribution of sample range and corresponding Pearson distribution, are given in Table 1.

When we compare the values of upper control limits calculated from theoretical distribution and corresponding Pearson distribution from Table 1, the approximation error ranges from 0.00049 (Weibull distribution) to 0.02636 (standard exponential distribution).

Now, we are interested to see what is the power of R control charts for detecting shifts $\nu = \dfrac{\sigma_1}{\sigma_0}$ from 2 to 6, for previously calculated upper control

Table 1. Upper control limits of R control chart

Sample size	**Upper control limit**					
	Exponential $E(1)$		*Gamma* $G(2,1)$		*Weibull* $W(2,1)$	
	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson
$n=3$	6.60698	6.59422	8.33514	8.32675	2.21175	2.20727
$n=4$	7.01222	6.99396	8.84503	8.82992	2.32690	2.32335
$n=5$	7.29978	7.27856	9.20631	9.18797	2.40740	2.40482
$n=6$	7.52285	7.49981	9.48556	9.46531	2.46880	2.46697
$n=7$	7.70514	7.68085	9.71276	9.69125	2.51818	2.51687
$n=8$	7.85926	7.83408	9.90403	9.88162	2.55930	2.55836
$n=9$	7.99273	7.96692	10.06904	10.04595	2.59443	2.59375
$n=10$	8.11053	8.08416	10.21402	10.19040	2.62501	2.62452

limits. Power of R control chart for detecting shifts from standard deviation σ_0 to $\sigma_1 = \nu\sigma_0$ can be calculated from

$$1-\beta = P\{R \geq UCL | \sigma = \sigma_1\} = 1 - F\left(\frac{UCL}{\nu}\right).$$

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

Calculated power of R control chart, for sample sizes from 3 to 10, shifts from 2 to 6, theoretical distribution of sample median and corresponding Pearson distribution, are given in Table 2.

From the Table 2, we see that R control charts can not detect shifts of sizes smaller than $\nu = 3$ with power of 90% and greater, for the chosen positively skewed distributions of quality characteristic. Even shift $\nu = 3$ is detected with power at least 90% only in the case of Weibull $W(1)$ distribution, for sample sizes of at least $n = 9$. In the case of standard exponential distribution of quality characteristic, shift has to be of size $\nu = 6$ and sample sizes $n \geq 9$ for R control chart to detect it with a power of at least 90%. In order to R control chart detect shifts of $\nu = 5$ with power of 90% and greater, for $G(2,1)$ distribution, it is necessary to take samples of sizes at least $n = 7$. Also, we can once more notice that Pearson distributions approximate well distribution of sample range. In general, it can be concluded that the R control chart can detect only very large shifts in standard deviation with power of 90% and greater, for positively skewed distributions of quality characteristic.

Table 2. Power of R control chart

Distribution	Power									
	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson	Theor.	Pearson
	$\nu = 2$		$\nu = 3$		$\nu = 4$		$\nu = 5$		$\nu = 6$	
$E(1)$										
$n = 3$	0.07216	0.07284	0.20887	0.21039	0.34668	0.34779	0.46236	0.46257	0.55442	0.55382
$n = 4$	0.08737	0.08857	0.26265	0.26507	0.43489	0.43609	0.57133	0.57088	0.67259	0.67115
$n = 5$	0.09999	0.10163	0.30744	0.31033	0.50503	0.50585	0.65255	0.65144	0.75468	0.75294
$n = 6$	0.11097	0.11297	0.34614	0.34923	0.56273	0.56305	0.71517	0.71365	0.81368	0.81216
$n = 7$	0.12078	0.12310	0.38032	0.38344	0.61123	0.61105	0.76449	0.76286	0.85702	0.85606
$n = 8$	0.12971	0.13230	0.41096	0.41400	0.65259	0.65200	0.80394	0.80244	0.88937	0.88913
$n = 9$	0.13794	0.14077	0.43874	0.44164	0.68826	0.68735	0.83587	0.83465	0.91382	0.91431
$n = 10$	0.14559	0.14865	0.46413	0.46684	0.71928	0.71815	0.86193	0.86112	0.93247	0.93364
$G(2,1)$										
$n = 3$	0.09997	0.10053	0.29039	0.29088	0.46330	0.46328	0.59263	0.59253	0.68530	0.68546
$n = 4$	0.12479	0.12605	0.37185	0.37327	0.58091	0.58093	0.72088	0.72038	0.80985	0.80957
$n = 5$	0.14590	0.14773	0.43893	0.44060	0.66873	0.66831	0.80613	0.80541	0.88347	0.88342
$n = 6$	0.16450	0.16680	0.49564	0.49718	0.73580	0.73500	0.86397	0.86348	0.92783	0.92839
$n = 7$	0.18127	0.18396	0.54437	0.54561	0.78780	0.78685	0.90378	0.90378	0.95492	0.95608
$n = 8$	0.19659	0.19961	0.58674	0.58763	0.82857	0.82769	0.93147	0.93204	0.97163	0.97320
$n = 9$	0.21074	0.21405	0.62392	0.62445	0.86083	0.86016	0.95095	0.95198	0.98225	0.98378
$n = 10$	0.22393	0.22748	0.65676	0.65697	0.88652	0.88618	0.96465	0.96611	0.98856	0.99029
$W(2,1)$										
$n = 3$	0.20599	0.20636	0.48959	0.48926	0.66747	0.66965	0.77144	0.77528	0.83485	0.83910
$n = 4$	0.27611	0.27662	0.62648	0.62713	0.80442	0.80665	0.88896	0.89129	0.93198	0.93368
$n = 5$	0.33760	0.33815	0.72586	0.72665	0.88480	0.88629	0.94602	0.94705	0.97197	0.97239
$n = 6$	0.39216	0.39269	0.79824	0.79884	0.93199	0.93286	0.97370	0.97410	0.98843	0.98845
$n = 7$	0.44092	0.44139	0.85111	0.85148	0.95974	0.96025	0.98715	0.98731	0.99521	0.99516
$n = 8$	0.48474	0.48511	0.88984	0.89005	0.97610	0.97645	0.99371	0.99379	0.99801	0.99798
$n = 9$	0.52431	0.52453	0.91830	0.91844	0.98578	0.98605	0.99691	0.99697	0.99917	0.99917
$n = 10$	0.56015	0.56021	0.93926	0.93940	0.99151	0.99175	0.99848	0.99853	0.99965	0.99966

3. Implementation of R control chart. Now we are interested to see how the R control chart can be implemented in practice, in general case when the distribution function of the quality characteristic is positively skewed but unknown. For fitting appropriate Pearson curve to data, we need estimates of mean, variance, skewness and kurtosis based on a sample of ranges. This procedure belongs to the Phase I of control chart usage, when we take preliminary (reference) samples in order to estimate unknown parameters of the distribution and construct control chart by calculating its center line and control limits.

Thus, we take N reference samples of size n and, for each sample, we calculate sample range. We wish to approximate the distribution of sample range by fitting Pearson's curve to data consisting of N sample ranges $R_1, R_2, \ldots, R_N$.

We will use $\bar{R}_N = \frac{1}{N}\sum_{i=1}^{N} R_i$ and $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(R_i - \bar{R}_N)^2$ as unbiased estimators of mean and variance of sample range. As for estimators of sample skewness and kurtosis, we shall investigate which of the various measures of skewness and kurtosis has minimum mean squared error (MSE), for non-normal samples. MSE is taken as a criterion for making a choice between estimators, as it incorporates both the variance of the estimator and its bias.

3.1. Choosing measures of sample skewness and kurtosis. Joanes and Gill [14] compared three available measures of sample skewness and excess kurtosis for normal and non-normal samples, on the basis of their bias and minimum squared error. Traditional measures of sample skewness and excess kurtosis g_1 and g_2 are equal to

$$g_1 = \frac{m_3}{m_2^{3/2}}, \quad g_2 = \frac{m_4}{m_2^2} - 3.$$

Sample moments for sample $(R_1, R_2, \ldots, R_N)$ are given by

$$m_k = \frac{1}{N}\sum_{i=1}^{N}(R_i - \bar{R}_N)^k.$$

In computing packages SAS, SPSS and EXCEL spreadsheet program, sample skewness and excess kurtosis are defined as

$$G_1 = \frac{\sqrt{N(N-1)}}{N-2} g_1, \quad G_2 = \frac{N-1}{(N-2)(N-3)}\left((N+1)g_2 + 6\right).$$

In MINITAB and BDMP, sample skewness and excess kurtosis are defined as

$$b_1 = \frac{m_3}{s^3}, \quad b_2 = \frac{m_4}{s^4} - 3.$$

Concerning the variance of these three measures of sample skewness and excess kurtosis, Joanes and Gill showed that the variances of G_1 and G_2 are greatest whereas b_1 and b_2 have the smallest variances, irrespective of the distribution being sampled. For non-normal samples, they generated samples of sizes 10, 20, 50 and 100 from χ^2_m-distributions with $m = 1, 10$ and 50 degrees of freedom (coefficients of skewness equal to 2.828, 0.894 and 0.4, respectively) and calculated bias and mean-squared error of these three measures. Their results showed that all the measures show negative bias with G_1 and G_2 less biased than g_1 and g_2, and b_1 and b_2 more biased than g_1 and g_2, respectively. The measures G_1 and G_2 have the smallest mean-squared error, and b_1 and b_2 have the largest, for the asymmetric χ^2_1-distribution (exception is the case of $n = 100$ where $mse(g_2) < mse(b_2) < mse(G_2)$). As for χ^2_{10} and χ^2_{50} distributions, measure of skewness b_1 has the smallest and G_1 largest mean-squared error. Further for $m = 10$ and $m = 50$, G_2 generally has the largest mean-squared error whatever the value of n. For small n, g_2 has the smallest mean-squared error whereas b_2 achieves the smallest mean-squared error for n sufficiently large, e.g. $n = 100$ for χ^2_{50}.

In order to investigate further the bias and mean squared error of measures of sample skewness and kurtosis, we will take, by Monte Carlo simulations, 100000 samples of sizes $N = 25, 50, 100$ from distributions of sample range of standard exponential, Gamma $G(2, 1)$ and Weibull $W(2, 1)$ distributions. Average estimates of bias and mean-squared error of these three measures are presented in Table 3. Note that because we are dealing with kurtosis throughout the paper, we added 3 to all the measures of excess kurtosis and denoted them with g_2^*, G_2^*, b_2^*, respectively. Also, we selected only sample sizes of 3, 5 and 10 from considered distributions of quality characteristic, to represent changes in values of skewness and kurtosis of distribution of sample range.

As it can be seen in Table 3, bias of these three measures of sample skewness and kurtosis is consistently negative. For $G(2, 1)$ and $W(1)$ distributions of quality characteristic, the bias of measures of sample skewness and kurtosis follows the relationship observed by Joanes and Gill.

In the case of standard exponential distribution, distribution of sample range is much more skewed (but not so highly skewed as χ^2_1 distribution). Measures g_1 and g_2^* are most biased of the measures of sample skewness and kurtosis, respectively. For some sample sizes and number of subgroups, G_1 and G_2^* are least biased, and for other values of n and N, b_1 and b_2^* are least biased.

Measure g_1 of sample skewness for range of standard exponential distribution has smallest mean-squared error, in most of the cases. As for ranges of

Table 3. Bias and mean squared error for skewness and kurtosis measures

Distribution	Skewness	Bias			Mean squared error			Kurtosis	Bias			Mean squared error		
		g_1	G_1	b_1	g_1	G_1	b_1		g_2^*	G_2^*	b_2^*	g_2^*	G_2^*	b_2^*
Exponential $E(1)$														
$n=3$ $N=25$	1.610	-0.470	-0.396	-0.396	0.577	0.559	0.556	7.080	-2.866	-2.296	-2.298	13.263	12.985	12.864
$n=3$ $N=50$		-0.283	-0.243	-0.240	0.383	0.384	0.385		-1.950	-1.593	-1.575	11.397	11.954	11.969
$n=3$ $N=100$		-0.162	-0.138	-0.141	0.258	0.260	0.260		-1.217	-0.993	-1.010	10.249	10.859	10.748
$n=5$ $N=25$	1.387	-0.406	-0.347	-0.344	0.503	0.498	0.502	6.194	-2.287	-1.810	-1.794	9.320	9.331	9.413
$n=5$ $N=50$		-0.244	-0.207	-0.207	0.338	0.342	0.338		-1.539	-1.222	-1.227	8.379	8.952	8.755
$n=5$ $N=100$		-0.139	-0.118	-0.118	0.225	0.224	0.226		-0.953	-0.770	-0.764	7.556	7.843	7.959
$n=10$ $N=25$	1.252	-0.373	-0.317	-0.314	0.470	0.474	0.475	5.738	-1.993	-1.545	-1.532	7.535	7.750	7.746
$n=10$ $N=50$		-0.217	-0.187	-0.190	0.312	0.317	0.317		-1.313	-1.036	-1.045	6.819	7.298	7.334
$n=10$ $N=100$		-0.127	-0.107	-0.108	0.204	0.208	0.210		-0.826	-0.645	-0.651	6.090	6.582	6.627
Gamma $G(2,1)$														
$n=3$ $N=25$	1.285	-0.349	-0.291	-0.403	0.426	0.431	0.430	5.576	-1.837	-1.391	-2.130	6.839	7.229	7.474
$n=3$ $N=50$		-0.205	-0.173	-0.236	0.282	0.281	0.281		-1.202	-0.931	-1.371	6.201	6.569	6.250
$n=3$ $N=100$		-0.114	-0.095	-0.132	0.179	0.181	0.179		-0.730	-0.562	-0.821	5.296	5.700	5.313
$n=5$ $N=25$	1.121	-0.314	-0.262	-0.362	0.402	0.406	0.397	5.131	-1.556	-1.154	-1.844	5.388	5.732	5.908
$n=5$ $N=50$		-0.184	-0.153	-0.212	0.260	0.269	0.260		-1.025	-0.756	-1.183	4.853	5.447	4.968
$n=5$ $N=100$		-0.103	-0.084	-0.117	0.167	0.169	0.164		-0.610	-0.455	-0.700	4.277	4.569	4.226
$n=10$ $N=25$	1.054	-0.301	-0.253	-0.349	0.391	0.405	0.390	4.997	-1.483	-1.078	-1.763	4.916	5.272	5.446
$n=10$ $N=50$		-0.175	-0.149	-0.203	0.263	0.264	0.255		-0.949	-0.718	-1.122	4.637	4.959	4.578
$n=10$ $N=100$		-0.099	-0.082	-0.114	0.164	0.167	0.161		-0.584	-0.431	-0.675	3.910	4.224	3.859
Weibull $W(1)$														
$n=3$ $N=25$	0.703	-0.147	-0.109	-0.180	0.212	0.224	0.200	3.440	-0.504	-0.233	-0.730	1.450	1.841	1.553
$n=3$ $N=50$		-0.077	-0.060	-0.097	0.120	0.124	0.118		-0.284	-0.144	-0.410	1.133	1.305	1.154
$n=3$ $N=100$		-0.042	-0.031	-0.050	0.067	0.068	0.066		-0.157	-0.078	-0.218	0.781	0.846	0.783
$n=5$ $N=25$	0.531	-0.119	-0.096	-0.146	0.213	0.235	0.197	3.331	-0.428	-0.172	-0.657	1.166	1.503	1.262
$n=5$ $N=50$		-0.064	-0.050	-0.080	0.125	0.129	0.119		-0.237	-0.098	-0.362	0.933	1.077	0.940
$n=5$ $N=100$		-0.033	-0.027	-0.041	0.068	0.069	0.066		-0.127	-0.057	-0.193	0.626	0.671	0.615
$n=10$ $N=25$	0.494	-0.121	-0.095	-0.143	0.230	0.255	0.212	3.420	-0.468	-0.189	-0.700	1.225	1.587	1.345
$n=10$ $N=50$		-0.066	-0.05	-0.080	0.135	0.143	0.130		-0.270	-0.115	-0.391	0.966	1.134	0.976
$n=10$ $N=100$		-0.035	-0.029	-0.042	0.076	0.077	0.074		-0.143	-0.071	-0.207	0.689	0.709	0.682

Gamma and Weibull distributions, b_1 has smallest squared-error (except for the case $n = 3, N = 25$ for $G(2,1)$ distribution when g_1 has smallest mean-squared error).

As for the measures of sample kurtosis, G_2^* has the largest mean-squared error whatever the value of n and N. In most of the cases, g_2^* has the smallest mean squared error.

Summing up Joanes and Gill's and our results we can give following recommendations concerning the choice of appropriate measures of sample skewness and kurtosis for positively skewed distributions: measures G_1 and G_2^* are appropriate for highly skewed distributions. In other cases, choice between b_1, b_2^* and g_1, g_2^* is not so clear, but our preferences go to g_1, g_2^*, as g_2^* has the smallest MSE in most of the cases and when MSE of b_1 is smaller than that of g_1, there is only a slight difference between their mean squared errors.

3.2. Example. Dou and Sa (2002) gave a data set on the viscosity of a certain chemical. When process was thought to be in control, 25 preliminary samples, each of size 10, were collected. By inspection of the histogram of the data (left graph on Figure 1), we can see that distribution of viscosity of a chemical is positively skewed. Non-normality of the data distribution is also confirmed by Shapiro-Wilk, Anderson-Darling and Cramer von Mises normality tests.

Average range of 25 reference samples is 6.9464, which represents the center line of R control chart. We calculated ranges of preliminary samples. All measures of sample skewness and excess kurtosis considered previously, can be found in R package $e1071$ (functions *skewness* and *kurtosis*). As the distribution of sample range is not highly skewed, we will use measures g_1 and g_2^*. We get $g_1 = 0.3439$, $g_2^* = 2.0640$. After calculations of coefficients c_0, c_1 and c_2 using formulas (4) and roots of quadratic equation (7), choice falls on Pearson I distribution for the approximation of a distribution of sample range. Parameters of corresponding Pearson distribution are calculated using formulas (9), given in the Appendix. On Figure 1, right graph, we constructed empirical cumulative distribution function along with fitted Pearson's curve. Now we may calculate upper control limit of R control chart using equation (1), for probability of false alarms $\alpha = 0.0027$. We get $k = 12.50673$. None of the ranges of reference samples are greater than the upper control limit, so we can conclude, in this stage, that process is in control and keep the calculated control limit. Dou and Sa collected 15 additional (test) samples from the same process. Ranges of these samples are represented on a graph in Figure 2. Vertical line at sample number 25 separates R control chart for reference and test samples. As we can see, the range of the 36 test sample falls out of the upper control limit, which requires the search for a possible assignable

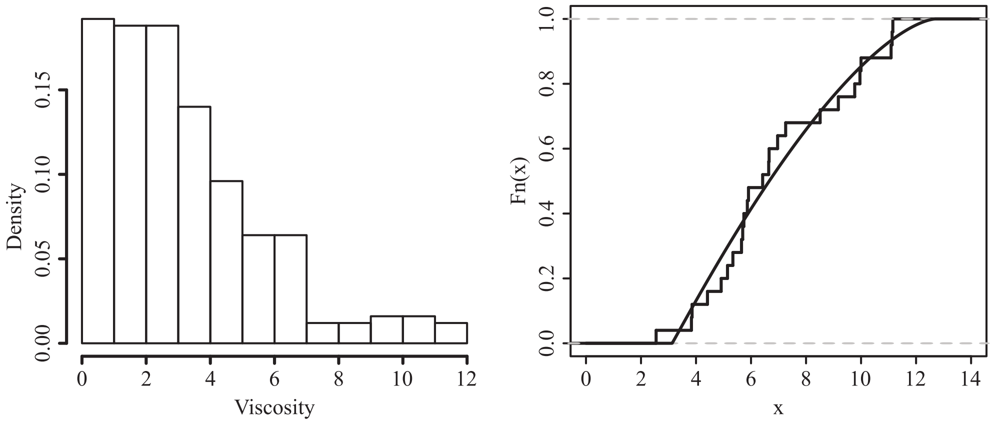

Fig. 1. Histogram of the viscosity data (left graph) and empirical cumulative distribution function of sample ranges with fitted Pearson I distribution(right graph)

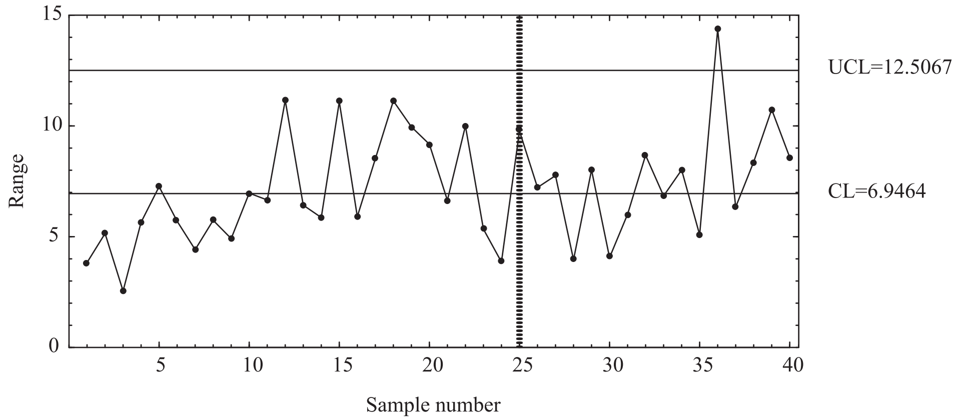

Fig. 2. R control chart for the viscosity data

cause to be initiated.

4. Conclusions. We considered design of the R control chart for positively skewed distributions. For chosen exponential, gamma and Weibull distributions of quality characteristic, we derived the theoretical distribution of sample range and approximated it with corresponding Pearson distribution. Then we calculated upper control limit of the R control chart, which gave evidence of the goodness of fit of the appropriate Pearson distribution to the theoretical distribution of sample range. Further, we examined the power of proposed chart in detecting shifts in process standard deviation. Results suggest that shifts have to be large for the R control chart to detect them with power of at least 90%. Finally, we considered implementation of proposed R chart in the general case, when distribution of quality characteristic is positively skewed but unknown. First, we made comparisons between measures of sample skewness and kurtosis necessary for fitting Pearson curve to data of sample ranges, based on the criterion of minimum mean squared error of an estimator. From Joanes and Gill's results, we concluded that measures G_1 and G_2^* are appropriate for highly skewed distributions. In other cases, choice between b_1, b_2^* and g_1, g_2^* is not so clear, but we preferred g_1, g_2^*, as g_2^* has the smallest MSE and MSE of b_1 is generally only slightly smaller than that of g_1. Finally, we constructed R control chart for a given data set.

Throughout the paper, we considered only Pearson type I and VI distributions for construction of R control chart. In many occasions, other Pearson distributions could be used for approximation of the distribution of sample range and this will be the subject of future research.

Appendix.

Distribution of quality characteristic. Distributions are given by their density function f.

Exponential distribution $E_2(\lambda, c)$

$$f(x) = \frac{1}{\lambda} e^{-\frac{x-c}{\lambda}}, \ x \geq c,$$

where λ ($\lambda > 0$) is a scale parameter, c ($c \in \mathbb{R}$) is a location parameter.

Gamma distribution $G_3(a, b, c)$

$$f(x) = \frac{1}{b\Gamma(a)} \left(\frac{x-c}{b}\right)^{a-1} e^{-\frac{x-c}{b}}, \ x \geq c,$$

where a ($a > 0$) is shape parameter, b ($b > 0$) scale parameter and c ($c \in \mathbb{R}$) location parameter.

Weibull distribution $W_3(a, b, c)$

$$f(x) = \frac{a}{b} \left(\frac{x-c}{b}\right)^{a-1} e^{-\left(\frac{x-c}{b}\right)^a}, \ x \geq c,$$

where a ($a > 0$) is shape parameter, b ($b > 0$) scale parameter and c ($c \in \mathbb{R}$) location parameter.

Theoretical distribution of sample range. Let $X_1, X_2, \ldots, X_n$ denote a random sample and $X_{(k)}$, $1 \leq k \leq n$, k-th order statistics. Sample range is defined as $R = X_{max} - X_{min}$, difference between the largest and smallest sample values.

Cumulative distribution function of sample range is equal to [11].

$$F_R(r) = n \int_0^{+\infty} f_X(x) \left(F(x+r) - F(x)\right)^{n-1} \mathrm{d}x, \; r > 0. \tag{2}$$

Pearson system of distributions. Pearson designed a system of distributions where probability density function $y = f(x)$ of each member satisfies a differential equation [15].

$$\frac{1}{y}\frac{dy}{dx} = -\frac{a+x}{c_0 + c_1 x + c_2 x^2}. \tag{3}$$

The shape of the distribution depends on the values of the parameters a, c_0, c_1, and c_2. The formulas for these coefficients are

$$c_0 = \frac{4\beta_2 - 3\beta_1}{10\beta_2 - 12\beta_1 - 18}\sigma^2 \tag{4}$$

$$a = c_1 = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18}\sigma \tag{5}$$

$$c_2 = \frac{2\beta_2 - 3\beta_1 - 6}{10\beta_2 - 12\beta_1 - 18}, \tag{6}$$

where $\beta_1 = \alpha_3^2$, $\beta_2 = \alpha_4$. Pearson classified the different shapes into a number of types. The form of solution of (3) depends on the nature of the roots of the equation

$$c_0 + c_1 x + c_2 x^2 = 0. \tag{7}$$

Pearson type I distribution. Pearson type I distribution corresponds to the case when both roots a_1 and a_2 of quadratic equation (7) are real and of opposite sign. We will denote $a_1 < 0 < a_2$.

Then, random variable $T = \dfrac{X - \mu - a_1}{a_2 - a_1}$ has Beta distribution $B(\alpha, \beta)$. Probability density function of random variable X is then equal to

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{1}{s} \left(\frac{x-\lambda}{s}\right)^{\alpha-1} \cdot \left(1 - \frac{x-\lambda}{s}\right)^{\beta-1}, \; 0 < \frac{x-\lambda}{s} < 1, \tag{8}$$

where

$$
\begin{aligned}
\lambda &= \mu + a_1, \quad s = a_2 - a_1, \\
\alpha &= \frac{a + a_1}{c_2(a_2 - a_1)} + 1, \quad \beta = -\frac{a + a_2}{c_2(a_2 - a_1)} + 1.
\end{aligned} \tag{9}
$$

Pearson type VI distribution. Pearson type VI distribution corresponds to the case when the roots a_1 and a_2 of quadratic equation (7) are real and of the same sign. We will consider the case when $a_1 < a_2 < 0$.

Then, random variable $T = \dfrac{X - \mu - a_2}{a_2 - a_1}$ has Beta prime distribution $B^{'}(\alpha, \beta)$. Probability density function of random variable X is then equal to

$$
f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{1}{s} \left(\frac{x-\lambda}{s}\right)^{\alpha-1} \cdot \left(1 + \frac{x-\lambda}{s}\right)^{-\alpha-\beta}, \quad \frac{x-\lambda}{s} > 0, \tag{10}
$$

where

$$
\begin{aligned}
\lambda &= \mu + a_2, \quad s = a_2 - a_1, \\
\alpha &= -\frac{a + a_2}{c_2(a_2 - a_1)} + 1, \quad \beta = \frac{1}{c_2} - 1.
\end{aligned} \tag{11}
$$

REFERENCES

[1] J. Alloway Jr., M. Raghavachari. Control charts based on Hodges-Lehmann estimator. *Journal of Quality Technology* **23**, *4* (1991), 336–347.

[2] L. C. Alwan. The problem of misplaced control limits. *J. Roy. Statist. Soc. Ser. C* **44**, *3* (1995), 269–278.

[3] D. S. Bai, I. S. Choi. $\bar{X}$ and R control charts for skewed populations. *J. Quality Technology* **27**, *2* (1995), 120–131.

[4] R. P. Brent. Algorithms for Minimization without Derivatives. New Jersey, Prentice-Hall, 1973.

[5] L. K. Chan, H. J. Cui. Skewness correction $\bar{X}$ and R charts for skewed distributions. *Naval Res. Logist.* **50**, *6* (2003), 555–573.

[6] Y. S. Chang, D. S. Bai. Control charts for positively-skewed populations with weighted standard deviations. *Quality and Reliability Engineering International* **17**, *5*, 2001, 397–406.

[7] F. Choobineh, J. L. Ballard. Control-Limits of QC charts for skewed distribution using weighted-variance. *IEEE Trans. on Reliability* **36**, *4* (1987), 473–477.

[8] Y. Dou, P. Sa. One-sided control charts for the mean of positively skewed distributions. *Total Quality Management* **13**, *7* (2002), 1021–1033.

[9] D. Djoric, J. Malisic, V. Jevremovic, E. Nikolic-Djoric. Atlas Raspodela. Gradjevinski fakultet, Beograd, 2007.

[10] E. B. Ferrel. Control charts for log-normal universes. *Industrial Quality Control* **15**, *2* (1958), 4–6.

[11] E. J. Gumbel. The Distribution of the Range. *The Annals of Mathematical Statistics* **18**, *3* (1947), 384–412.

[12] H. L. Harter. Tables of range and Studentized range. *The Annals of Mathematical Statistics* **31**, *4* (1960), 1122–1147.

[13] G. J. Janacek S. E. Meikle. Control charts based on medians. *The Statistician* **46**, *1* (1997), 19–31.

[14] D. N. Joanes, C. A. Gill. Comparing measures of sample skewness and kurtosis. *The Statistician* **47**, *1* (1998), 183–189.

[15] N. L. Johnson, S. Kotz, N. Balakrishnan. Continuous Univariate Distributions Volume 1. New York, Wiley, 1994.

[16] N. L. Johnson, S. Kotz, N. Balakrishnan. Continuous Univariate Distributions Volume 2. New York, Wiley, 1995.

[17] S.-C. Kao, C. Ho. Robustness of R-Chart to Non Normality. *Communications in Statistics – Simulation and Computation* **36**, *5* (2007), 1089–1098.

[18] J. F. Mahoney. The influence of parent population distribution on d2 values. *IIE Transactions* **30**, *6* (1998), 563–569.

[19] D. C. Montgomery. Introduction to Statistical Quality Control. New York, Wiley, 2005.

[20] P. R. Nelson. Control charts for Weibull Process with standards given. *IEEE Transaction on Reliability* **28**, *4* (1979), 283–288.

[21] N. Pal, J. Chun, K. L. Wooi. Handbook of Exponential and Related Distributions for Engineers and Scientists. Chapman and Hall/CRC, 2005.

[22] E. S. Pearson, H. O. Hartley. The probability integral of the range in samples of n observations from a normal population. I. Foreword and tables. II. Numerical evaluation of the probability integral. *Biometrika* **32**, *3/4* (1942), 301–308, 309–310.

[23] H. Rinne. The Weibull distribution. A Handbook. Boca Raton, FL, CRC Press, 2009.

[24] L. H. C. Tippett. On the extreme individuals and the range of samples taken from a normal population. *Biometrika* **17**, *3/4* (1925), 364–386.

[25] T.-R. Tsai, S.-J. Wu. Adjusted weighted standard deviation R chart for skewed distributions. *Braz. J. Probab. Stat.* **22**, *1* (2008), 9–22.

Department of Probability and Statistics
Faculty of Mathematics
Belgrade, Serbia
e-mail: `kristina@matf.bg.ac.rs`

Received October 15, 2015
Revised March 16, 2016