IJST (2014) 38A3: 289-294
Iranian Journal of Science & Technology

http://ijsts.shirazu.ac.ir

Approximation effects on control charts for process monitoring

M. Riaz

Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia
E-mail: riazqau76@yahoo.com

Abstract

In Statistical quality control a very widely used measure is average run length (ARL) which may be worked out by different methods like integral equation, approximations, and Monte Carlo simulations. The ARL measure and the other related measures are of major significance in every type of production process. An omission in its computation (and hence its related measures such as extra quadratic loss (EQL)) may cause a loss. This necessitates great care in the choice of its evaluation method. This article will deal with this issue using some approximation methods and the Monte Carlo simulations. The discrepancies among the results will be examined to highlight the deficiencies of using approximation methods in quality control techniques.

***Keywords**:* Approximations; average run length (arl); process parameters; statistical process control (spc); simulations; $\bar{X}$ - chart

1. Introduction

In Statistical Process Control (SPC) we are looking for an efficient monitoring technique which may ensure stability of the parameters of interest in a process. These parameters may be location, scale, association, proportion, etc. and quality of the process output depends on how timely we are able to detect any change in the behavior of these process parameters. There are a number of tools which may be used for this purpose like Pareto chart, Cause and Effect diagram, Check sheet, etc. The kit comprising these tools is formally known as SPC tool-kit which is used for detecting shifts in process behavior in terms of its parameters.

Monitoring of the process parameters is one key to enhancing the performance of the output. For this purpose many statistical tools are used in practice and control chart is one of the most popular choices. The most commonly used control charts for location and spread parameters are $\bar{X}$, R and S charts, proposed by Shewhart back in the 1920s. There are a variety of classifications to define the design structure of a control chart such as parametric versus non-parametric, variable versus attributes, univariate versus multivariate, bayesian versus classical, simple versus ranked set sampling etc. In process monitoring using control charts an important performance measure is Average Run Length (ARL), along with some other useful measures like standard deviation run length (SDRL), relative ARL (RARL), extra quadratic loss (EQL), performance comparison index (PCI). The details about the said measures may be seen in Zhang and Wu (2006); Wu et al., (2009); Ou et al., (2011); Ou et al., (2012); Ahmad et al., (2013) and the references therein.

The ARL is interpreted as the expected number of samples we have to wait to receive an out-of-control signal. Now a distortion in this number consequently may affect other performance measures such as SDRL, RARL, EQL and PCI, which could ultimately harm the process outcomes. The ARL measure is mathematically defined as an integral which is sometimes analytically not solvable. In such situations different authors have various approaches to work out this number including approximations and simulations. Reynolds (1975); Champ and Rigdon (1991); Hawkins (1992); Fu et al., (2002); Schaffer and Kim (2007); Hanif et al., (2012) and the references therein may be seen for some relevant literature on it.

Champ and Rigdon (1991) mentioned that Markov chain and integral equation approaches are often used to evaluate run length measures for control charts. They indicated that Markov chain approach uses discretization of the possible values in order to reach at the run length properties. The integral equation approach depends on the evaluation of the integral analytically or using some approximation methods. They noted that under certain conditions (such as using midpoint rule in

Received: 12 July 2013 / Accepted: 16 April 2014

their case) the two approaches lead to the sane approximation results. Therefore for our study purposes we have not included Markov chain approach and used the other approximation methods.

In order to compute these measures like ARL, SDRL and the other related measures, we need to apply them very carefully. Any skipping or omission of these approximations in quality control applications (where we mainly deal with smaller sample sizes) may distort these performance measures. This causes loss of efficiency in the production processes. In this study we will consider the approximation methods of Menzefricke (2002) and the same will be evaluated here through extensive Monte Carlo Simulations. The discrepancies among the results of all different approaches will be highlighted in the next section.

2. Average Run Length Evaluation using Different Techniques

In Menzefricke (2002) a control chart is proposed for the location parameter of a normally distributed process using Bayesian approach. He considered prior uncertainty in the form of prior distribution, the characteristic of interest in the form of sampling distribution, the updated information in the form of posterior distribution and finally used posterior predictive distribution to develop the design structure of his proposed Bayesian $\bar{X}$ -Chart. He assumed that: the characteristic of interest $X \sim N\left(\mu,\sigma^2\right)$ where μ and σ are location and scale parameters respectively; the prior uncertainty of $\mu \sim N\left(m_0,\sigma^2/n_0\right)$ where m_0 is a priori estimate of μ and n_0 is the prior sample size. This implies that for the current sample $x_1,x_2,...,x_{n_c}$ of size n_c we have: the current sample mean $\bar{x} \sim N\left(\mu,\sigma^2/n_c\right)$; the posterior distribution $\mu/\bar{x} \sim N\left((n_0m_0+n_c\bar{x})/(n_0+n_c),\sigma^2/(n_0+n_c)\right)$. Consequently for the future sample $y_1,y_2,...,y_n$ of size n we have: the future sample mean $\bar{y} \sim N\left(\mu,\sigma^2/n\right)$; the posterior predictive distribution $\bar{y}/\bar{x} \sim N\left(\bar{x}+(n_0/(n_0+n_c))(m_0-\bar{x}),\,(\sigma^2/n)(1+n/(n_0+n_c))\right)$.

Based on the above mentioned setup, Menzefricke (2002) defined the Bayesian control limits (specified parameters) as $\mu_0+(n_0/(n_0+n_c))(m_0-\mu_0)\pm Z_{1-\alpha/2}(\sigma/\sqrt{n})\sqrt{1+n/(n_0+n_c)}$ using predictive distribution and calculated their performance with the help of the integral $\int \frac{\Phi(z)}{(1-\Phi(u_2-\sqrt{n/n_1}z)+\Phi(u_1-\sqrt{n/n_1}\,z))}dz$, where $\Phi()$ denotes the standard normal density, $\Phi(z)$ denotes the standard normal distribution function evaluated at z, $u_1=Z_{\alpha/2}\sqrt{1+n/n_1}$, $u_2=Z_{1-\alpha/2}\sqrt{1+n/n_1}$ and $n_1=n_0+n_c$. He mentioned that these measures may be obtained using one dimensional numerical integration. He provided the results for mean and standard deviation of the in-control distribution of Run Length (RL) (i.e. ARL_0 and SDR_0) (we will refer to these results of Menzefricke (2002) by NI method in this study). An alternative he suggested in the form of an approximation for ARL_0 is given by $[2\Phi(u_1)\sqrt{1-(nu_1\Phi(u_1)/n_1\Phi(u_1))}]^{-1}$ using Laplace's method (we will refer to these results by APP approach in this study).

The above mentioned ARL and SDRL properties are functions of the ratio n/n_1 and Menzefricke (2002) evaluated their values for certain choices of n/n_1 using NI and APP. The same results have been evaluated here using Monte Carlo simulation approach (we will refer to these results by MCS in this study). The results of ARL_0 and $SDRL_0$ by the said approaches are given in Table 1 for different choices of n/n_1. For Monte Carlo simulations $5\text{x}10^5$ repetitions are used, which may be increased to improve the precision of the results further. Schaffer and Kim (2007) may be seen for discussion on the number of replications needed in control chart Monte Carlo simulation studies,

The errors (in percentage terms) for different techniques including NI, APP and MCS are also reported for ARL_0 and $SDRL_0$ in Table 1. The off-centering effect is considered almost negligible, however, if present it may affect the properties (see Riaz, 2011). The differences in ARL_0 and $SDRL_0$ among different techniques and their corresponding error rates are also shown in the form of graphs shown in Figs. 1-3 for ease in comparison.

Table 1. ARL_0 and $SDRL_0$ Values for Three Methods and their Corresponding Error Rates

	ARLo			Error Rates (%) for ARLo			$SDRL_0$		Error Rates (%) for $SDRL_0$
n/n_1	APP	NI	MCS	MCS vs. NI (ARLo)	MCS vs. APP (ARLo)	NI vs. APP (ARLo)	NI	MCS	MCS vs. NI (SDRL)
0	370	370	370.16	0.043	0.043	0.000	370	369.98	0.007
0.01	371	389	390.74	0.446	5.053	4.627	389	388.86	0.036
0.02	373	408	408.54	0.133	8.700	8.578	407	409.68	0.655
0.05	385	468	474.79	1.429	18.911	17.735	467	476.52	1.998
0.1	420	576	604.89	4.776	30.566	27.083	578	597.65	3.288
0.2	540	832	981.89	15.266	45.004	35.096	802	974.98	17.742
0.5	1464	2242	4237.50	47.091	65.451	34.701	2925	4260.62	31.348
1	10143	12677	45557.00	72.173	77.736	19.989	24231	44766.83	45.873
0	370	370	370.16	0.043	0.043	0.000	370	369.98	0.007

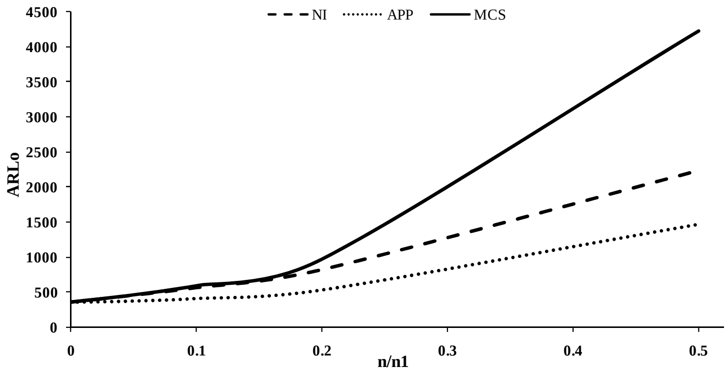

Fig. 1. ARL0 Curves for Different Techniques

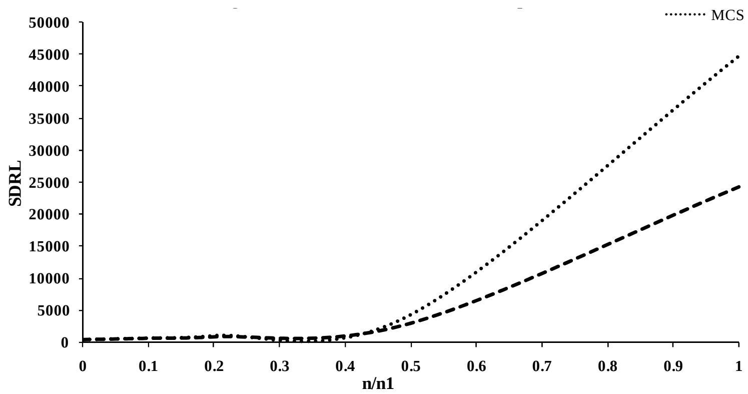

Fig. 3. Comparative Error Rates of ARLo and SDRL for Different Techniques

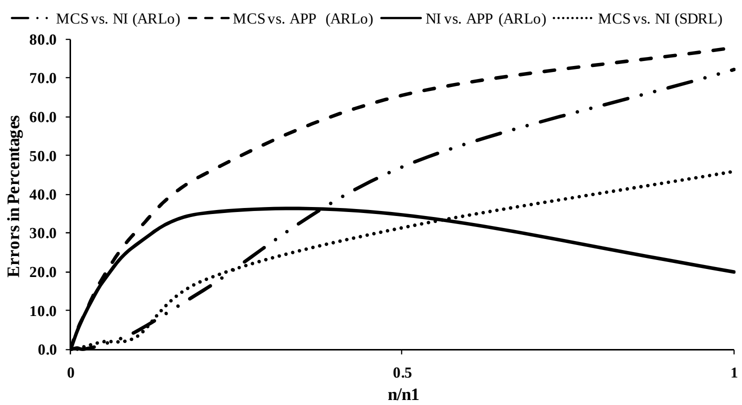

Fig. 3. Comparative Error Rates of ARLo and SDRL for Different Techniques

From Table 1 and Figs. 1-3 it is advocated that:
i) APP causes the most distortion in the results for all choices of n/n_1 as obvious from Table 1 and Fig. 1.
ii) NI works reasonably well for smaller choices of n/n_1 as may be seen in Table 1 and Figs. 1 and 2.
iii) MCS offers better computations of run length properties for all choices of n/n_1 as apparent from Table 1 and Figs. 1 and 2. Note that in Figs. 1 and 2 the gaps among different curves indicate the amount of computational difference among different techniques.
iv) the error rates between different approaches keep increasing with an increase in the ratio n/n_1, except for NI versus APP, as can be observed from Table 1 and Fig. 3.

The above points (i)-(iv) indicate that NI and APP methods introduce an error in the in-control run length properties with an increase in the ratio n/n_1. It would eventually seriously affect the out-of-control run length properties as well. In reality the behavior of the integrand does not remain constant over the complete range of the fraction n/n_1. This causes a varying amount of error over the changing values of n/n_1, especially for the cases where the ratio n/n_1 gets larger. It is more likely to come across larger values of the ratio n/n_1 in practical situations to exploit the Bayesian environment in its true spirit and get more benefit out of prior information available (cf. Berger, 1985; Bernardo and Smith, 1994). Therefore, the application of these computational techniques needs attention to get better run length properties both in the in-control and out-of control process situations.

The out-of-control run length study is also carried out by considering different amounts of shifts. The shifts are considered in terms of δ defined as: $\delta = (\mu_1 - \mu_0)/\sigma$ where μ_1 is shifted mean, μ_0 is in-control mean and δ=0 implies in-control otherwise out-of-control. We have considered varying choices of n/n_1. The resulting ARL and SDRL outcomes are reported in Tables 2 and 3 respectively, for n=4 as an illustration (for other values of n one may observe the deviance in results on similar lines). Table 2 contains some additional measures, along with ARL, for overall performance including RARL, EQL and PCI (where n/n_1=0 is used as benchmark). The results clearly indicate the distortion in different measures, especially for the larger values of n/n_1.

3. Concluding Remarks

In SPC we are mainly dependent on the run length properties to investigate the process behavior and its performance. Any skip in the computation of these properties may cause a loss in the final outcome of the process due to misleading run length results. This necessitates great care in the choice of the evaluation method used for exploring run length properties. It has been examined in this study that different computational/approximation methods depend very much on the nature of the function under investigation and give an under estimation in run length values in general due to omission of important terms in the evaluation of integral. Therefore, it is important to guard against these types of skips for more accurate run length properties in both the in-control and out-of-control process environments. Moreover, in case of mixture behaviors over the varying values of the parameter quantities (like n/n_1) we may opt a merger of different computation methods depending upon the changing values of the parameter quantities.

Table 2. ARL Values under Varying Choices of n/η_1

δ \ n/η_1	**0**	**0.01**	**0.02**	**0.05**	**0.1**	**0.2**	**0.5**	**1**
-3	1.001	1.001	1.002	1.002	1.003	1.004	1.01	1.046
-2.5	1.024	1.025	1.025	1.027	1.032	1.044	1.107	1.274
-2	1.187	1.196	1.199	1.213	1.245	1.315	1.587	2.51
-1.5	2.005	2.025	2.052	2.127	2.224	2.586	4	9.276
-1	6.294	6.45	6.602	7.043	7.987	10.225	21.403	78.819
-0.75	14.989	15.418	15.854	17.193	20.105	27.034	65.967	333.106
-0.5	43.843	45.269	47.199	52.584	62.97	88.983	269.364	1705.717
-0.25	155.011	162.205	169.193	191.093	240.807	364.441	1299.481	10252.41
0	370.16	390.743	408.544	474.786	604.889	981.894	4237.496	45557
0.25	155.017	161.574	168.377	191.794	240.897	365.323	1302.845	10610.8
0.5	44.003	45.543	47.063	52.427	62.73	88.856	268.491	1728.063
0.75	14.994	15.41	15.885	17.301	20.144	27.071	68.531	318.93
1	6.332	6.452	6.595	7.097	7.94	9.987	21.273	78.177
1.25	3.255	3.292	3.358	3.51	3.828	4.672	8.465	24.477
1.5	1.999	2.027	2.046	2.12	2.259	2.609	4.029	9.438
2	1.19	1.195	1.2	1.219	1.248	1.311	1.588	2.378
2.5	1.024	1.023	1.026	1.027	1.035	1.046	1.097	1.287
3	1.001	1.002	1.002	1.002	1.002	1.003	1.012	1.048
EQL	6.561	6.677	6.803	7.193	7.980	9.949	22.588	121.974
RARL	1.000	1.017	1.034	1.089	1.199	1.480	3.300	18.735
PCI	1.000	1.018	1.037	1.096	1.216	1.516	3.443	18.591

Table 3. SDRL Values under Different Out-of-Control Situations

δ \ n/η_1	**0**	**0.01**	**0.02**	**0.05**	**0.1**	**0.2**	**0.5**	**1**
-3	0.036	0.035	0.042	0.04	0.058	0.067	0.099	0.21
-2.5	0.156	0.158	0.161	0.166	0.183	0.217	0.345	0.574
-2	0.471	0.482	0.487	0.503	0.557	0.64	0.947	1.994
-1.5	1.418	1.437	1.464	1.545	1.654	2.038	3.418	8.575
-1	5.786	5.933	6.064	6.535	7.56	9.556	20.879	77.524
-0.75	14.504	14.981	15.405	16.647	19.689	26.454	65.421	329.911
-0.5	43.28	44.763	46.355	51.996	63.928	89.505	263.555	1651.199
-0.25	153.964	161.949	168.352	190.932	243.338	368.038	1299.975	9988.535
0	368.475	388.86	409.684	476.521	597.651	974.98	4260.624	44766.83
0.25	154.716	161.849	168.228	190.077	243.775	356.698	1292.028	10265.54
0.5	43.585	44.73	46.42	52.117	62.135	87.682	265.887	1723.09
0.75	14.492	14.913	15.477	16.723	19.891	26.418	66.717	309.07
1	5.779	5.905	6.062	6.575	7.321	9.342	20.929	76.99
1.25	2.702	2.733	2.807	2.967	3.246	4.164	8.168	24.31
1.5	1.416	1.444	1.461	1.542	1.688	2.055	3.516	8.935
2	0.475	0.481	0.49	0.516	0.569	0.641	0.966	1.725
2.5	0.154	0.154	0.164	0.166	0.189	0.217	0.319	0.617
3	0.037	0.039	0.04	0.043	0.044	0.056	0.112	0.232

Acknowledgements

The author is grateful to the editor and referees for their constructive comments that led to substantial improvements in the article. The author is also indebted to the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, for providing excellent research facilities.

References

Ahmad, S., Riaz, M., Abbasi, S. A., & Lin, Z. (2013). On Monitoring Process Variability under Double Sampling Scheme. *International Journal of Production Economics*, 142(2), 388–400.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, Berlin, Heidelberg, London.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian Theory*, John Wiley and Sons Ltd. Chichester, England.

Champ, C. W., & Rigdon, S. E. (1991). A Comparison of the Markov Chain and the Integral Equation Approaches for Evaluating the Run Length Distribution of Quality Control Charts. *Communications in Statistics: Simulation and Computation*, 20(1), 191–204.

Fu, J. C., Spiring, F. A., & Xie, H. (2002). On the average run lengths of quality control schemes using a Markov chain approach. *Statistics & Probability letters*, 56(4), 369–380.

Hanif, M., Hussain, A., Jamal, N., & Amir, M. (2012). New approximation of ARL in CUSUM control chart. *Far East Journal of Marketing and Management*, 2(2), 73–81.

Hawkins, D. (1992). A fast accurate approximation for average run lengths of CUSUM control charts. *Journal of Quality Technology*, 24(1), 37–43.

Menzefricke, U. (2002). On the evaluation of control chart limits based on predictive distributions. *Communication Statistics-Theory and Methods*, 31(8), 1423–1440.

Ou, Y. J., Wu, Z., & Goh, T. N. (2011). A new SPRT chart for monitoring process mean and variance. *International Journal of Production Economics*, 132(2), 303–314.

Ou, Y. J., Wu, Z., & Tsung, F. (2012). A comparison study of effectiveness and robustness of control charts for monitoring process mean. *International Journal of Production Economics*, 135(1), 479–490.

Reynolds, M. R. (1975). Approximations to the average run length in cumulative sum control charts. *Technometrics*, 17(1), 65–71.

Riaz, M. (2011). Monitoring of Process Parameters Using Bayesian Methodology. *International Journal of Agricultural and Statistical Sciences*, 7(1). 1–7.

Schaffer, J. R., & Kim, M. J. (2007). Number of replications in control chart Monte Carlo simulation studies. *Communications in Statistics-Simulation & Computation*, 36(5), 1075–1087.

Wu, Z., Jiao, J. X., Yang, M., Liu, Y., & Wang, Z. J. (2009). An enhanced adaptive CUSUM control chart. *IIE Transactions*, 41(7), 642–653.

Zhang, S., & Wu, Z. (2006). Monitoring the process mean and variance by the WLC scheme with variable sampling intervals. *IIE Transactions*, 38(4), 377–387.