Advanced Studies in Contemporary Mathematics
30 (2020), No. 4, pp. 571 – 584

www.jangjeon.or.kr
http://dx.doi.org/10.17777/ascm2020.30.4.571

DISTRIBUTION OF RUN-LENGTH FOR MULTIVARIATE CONTROL CHARTS WITH RULES USING EMBEDDED FINITE MARKOV CHAIN

HAN-BYUL LEE AND GYO-YOUNG CHO

ABSTRACT. A multivariate control chart that observes various characteristics at the same time was proposed by Hotelling. However, Hotelling's control chart has the disadvantage of slowing the speed of detecting anomalies when very small changes in the process occur, so rules for multivariate control charts have been proposed. In this paper, to increase the sensitivity of the Shewhart chart, which has the same shortcomings, the proposed rule, Western Electronic Company's rule, is applied to the multivariate chart to obtain the run-length distribution. The run-length distribution is calculated by recognizing that an abnormality occurs in the process when the mean vector of the process changes in the production process. The run-length distribution of the multivariate control chart obtained using the proposed runs rule and the embedded finite Markov chain confirmed that the mean of run-lengths decreased sharply when very small changes occurred in the process. This means the speed of detecting small changes in the process is fast. Therefore, this method is expected to be effective when observing very small changes due to multivariate quality control problems.

2010 MATHEMATICS SUBJECT CLASSIFICATION. 11T23, 20G40, 94B05.

KEYWORDS AND PHRASES. ARL, finite Markov chain embedding, mean control chart, runs rules, run-length distribution.

1. INTRODUCTION

In 1924, Shewhart confirmed that certain fluctuations were observed by measuring the quality of the product for a certain period of time, and proposed a control charts to find out the regularity of this change and to make an objective judgment by applying a simple statistical method. Then he first used the term quality control in *Economic control of Quality of Manufactured Products*(1931). Control charts can be applied wherever there is dispersion of data, so they can be used in a wide variety of ways. The Shewhart control charts are not susceptible to gradual process changes or small changes in process. To compensate for this, the CUSUM(Page, 1954) control charts and the EWMA(Roberts, 1959) control charts have been developed and are responded to small changes in the process more sensitively than the control charts. However, there are difficulties in using it with difficult theories and complex systems. We apply runs rules to the control charts to respond sensitively to slow and small changes. Runs rules for improving the sensitivity of the Shewhart control charts have been proposed continually by the Western Electronic Company(1956), Bissell(1978), Nelson(1999), among others. Champ and Woodall(1987) added the british limits to the first proposed rules by the Western Electronic Company(1956) in the above runs rules, and used the Markov chain to obtain the accurate average run-length(ARL) with some combined rules. In addition, Shmueli and Cohen(2003) proposed a run-length generating function as a way to derive an accurate run-length distribution with the above rules. We also want to extend this to multivariate problems. Two or more related quality characteristics monitoring or control problems are referred in the literature as multivariate quality control problems and the work was initiated by Hotelling(1947). The most widely used multivariate control charts for mean vector monitoring is Hotelling's χ^2 control chart, but it is not sensitive to the detection of small and moderate shifts in the mean vector. To make up for this, the rule is proposed by Khoo and Quah(2003). We will try to derive the ARL and run-length distribution of the Hotelling's χ^2 control charts using the rules proposed by the Western Electronic Company.

In this paper, we introduce the finite Markov chain embedding method and explains the runs rules related to multivariate and the ARL and run-length distribution for each runs rules. Finally, we will display the results of ARL, quartile values, cumulative distribution and probability distribution of run-length in our proposed method.

2. Finite Markov Chain embedding

2.1. **Finite Markov Chain embedding.** The method of applying the finite markov chain embedding technique to random variable $X_n(\Lambda)$ was introduced by Fu and Koutras(1994). Let $\Gamma_n = \{1, 2, \cdots, n\}$ be an index set and $\Omega = \{a_1, \cdots, a_m\}$ be a finite state space. If a non-negative integer random variable $X_n(\Lambda)$ satisfies the following conditions, it is possible to embed the finite markov chain. First, for a finite state space Ω , there exists a finite markov chain $\{Y_t : t \in \Gamma_n\}$. As the second condition, a finite partition $\{C_x, x = 1, 2, \cdots, l_n\}$ exists on the finite state space Ω. And for every $x = 0, 1, \cdots, l_n$, we have

$$P(X_n(\Lambda) = x) = P(Y_n \in C_x|\boldsymbol{\xi}_0), \tag{1}$$

where $\boldsymbol{\xi}_0$ is the initial probability distribution on the state space.

Fu and Koutras(1994) suggested to calculate the probability through the following theorem when the random variable $X_n(\Lambda)$ is embeddable the finite markov chain.

$$P(X_n(\Lambda) = x) = \boldsymbol{\xi}_0(\prod_{t=1}^{n} \boldsymbol{M}_t)\boldsymbol{U}'(C_x), \tag{2}$$

where $\boldsymbol{M}_t$, $t = 1, \cdots, n$ are the transition probability matrices of the embedded Markov Chain and $\boldsymbol{U}'(C_x) = \sum_{r:a_r \in C_x} \boldsymbol{e}_r$, $\boldsymbol{e}_r$ is a $1 \times m$ unit vector which coincides r-th state of the state space Ω.

2.2. **Waiting-Time Distribution.** Waiting time distributions of runs and patterns for Bernoulli trials such as geometric, geometric of order k, negative binomial *etc*, have been studied by Aki(1985), Feller(1968), Koutras(1996), Philippou(1986) and others.

Given a simple pattern $\Lambda = b_{i_1} \cdots b_{i_k}$, the number of trials until the first occurrence of a simple pattern appears is called the waiting time random variable $W(\Lambda)$ *i.e.*

(3) $$W(\Lambda) = inf\{n : n \geq k, X_{n-k+1} = b_{i_1}, \cdots, x_n = b_{i_k}\}.$$

For a compound pattern $\Lambda = \bigcup_{i=1}^{l} \Lambda_i$, the waiting time random variable $W(\Lambda)$ is

(4)
$$W(\Lambda) = inf\{n : occurrence\ if\ any\ pattern\ \Lambda_1, \cdots, \Lambda_l\ at\ the\ n-th\ trial\}.$$

Some theorem of the waiting time distribution is derived details and results using the embedded markov chain by Fu and Lou(2003).

The waiting time random variable $W(\Lambda)$ is associated with the state space Ω of the embedded markov chain $\{Y_t\}$. Let $\mathcal{S}$ and $\mathcal{S}(\Lambda)$ are the set of possible outcome $\{b_1, \cdots, b_k\}$ and sub-patterns of $\Lambda = \bigcup_{i=1}^{l} \Lambda_i$ and $\emptyset$ is a dummy state with $P(Y_0 = \emptyset) = 1$ as the initial state. Then the state space

$$\Omega = \{\emptyset\} \cup \{\mathcal{S}\} \cup \{\mathcal{S}(\Lambda)\} = \{\emptyset\} \cup \{\mathcal{S}\} \cup \{\mathcal{S}(\Lambda_1)\} \cup \cdots \cup \{\mathcal{S}(\Lambda_l)\},$$

and we can define a transition probability matrix of the from

(5) $$\boldsymbol{M} = \begin{matrix} \Omega - A \\ A \end{matrix} \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{O} & \boldsymbol{I} \end{array} \right),$$

where $A = \bigcup_{i=1}^{l} \Lambda_i$ is an absorbing state which means that the one of simple patterns has first occurred before the n-th trial.

Using above the matrices $\boldsymbol{N}$, the waiting time distribution for l simple patterns with the initial distribution $\boldsymbol{\xi}$ satisfied $P(Y_0 = \emptyset) = 1$ and a row vector $\mathbf{1}$ is given by

(6) $$P(W(\Lambda) = n) = \boldsymbol{\xi N}^{n-1}(\boldsymbol{I} - \boldsymbol{N})\mathbf{1}',$$

or

(7) $$P(W(\Lambda) \geq n) = \boldsymbol{\xi N}^{n-1}\mathbf{1}'.$$

Also, mean of waiting time $W(\Lambda)$ is derived by the theorem.

(8) $$EW(\Lambda) = \sum_{n=1}^{\infty} \boldsymbol{\xi N}^{n-1}\mathbf{1}' = \boldsymbol{\xi}(\boldsymbol{I} - \boldsymbol{N})\mathbf{1}'.$$

3. Runs Rules and Distribution of Run-Length

3.1. **Runs Rules.** First we will introduce runs and scans by Balakrichnan and Koutras(2002). The concept of run is an uninterrupted sequence. In other words, a run is an execution in which a test is repeated until a particular result is continuously obtained in a test in which a mutually exclusive event occurs. Considering a sequence of length m, if the sequence consists k successes (or failures) and k is close to m then the sequence is called a scan. Run and scan are used in the next runs rules.

Suppose a random variable X that measures the quality of a product and has a normal distribution with known μ_0 and σ_0. We can observe independent $\bar{X}_1, \bar{X}_2, \cdots$ and all $\bar{X}_i \sim N(\mu_i, \sigma_0^2/n), i = 1, 2, \cdots$. Then $Z_i =$

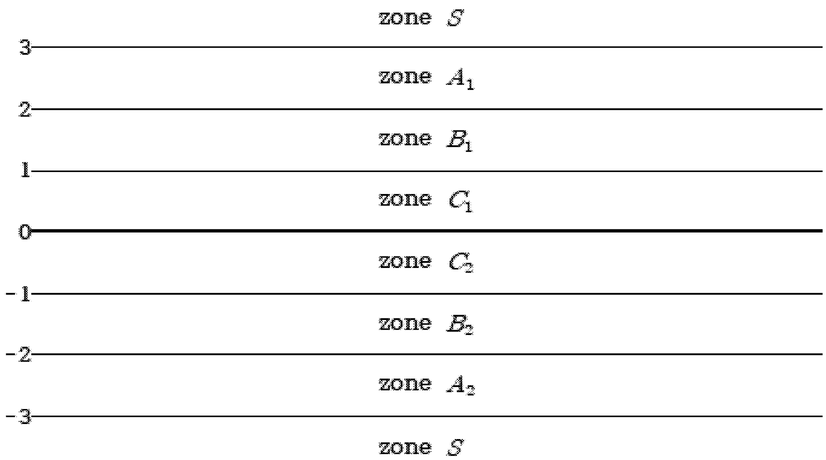

FIGURE 1. Seven zones for control chart

$(\bar{X}_i - \mu_0)/(\sigma_0/\sqrt{n})$ has a standard normal distribution and it signals if $|Z_i|$ is greater than c, where usually $c = 3$, on Shewhart control chart. However this procedure is not sensitive to detect small shift of the mean in the production process. We can improve the sensitivity of the charts by adding rules that tell us that process is out of control. We introduce the runs rules and how to use it.

By using the Champ and Woodall(1987)'s notation $T(k, m, Z)$, we describe the runs rules which signals if k of the last m standardized sample means fall in the zone Z. The zone is divided seven zones as shown in Figure 1 of a standard normal distribution.

Following rules are considered in this paper.

Rule 1 : $R_1 = T(1, 1, S)$
Rule 2 : $R_2 = T(2, 3, A_1) \cup T(2, 3, A_2)$
Rule 3 : $R_3 = T(4, 5, A_1 \cup B_1) \cup T(4, 5, A_2 \cup B_2)$
Rule 4 : $R_4 = T(8, 8, A_1 \cup B_1 \cup C_1) \cup T(8, 8, A_2 \cup B_2 \cup C_2)$
Rule 5 : $R_5 = T(2, 2, A_1) \cup T(2, 2, A_2)$
Rule 6 : $R_6 = T(5, 5, A_1 \cup B_1) \cup T(5, 5, A_2 \cup B_2)$

3.2. **Runs Rules for Multivariate Control Chart.** We defines the zones for applying the rules of Section 3.1 to the multivariate distribution. Assume that a product in which p different characteristics $x_1, x_2, \cdots, x_k$ are measured simultaneously in a certain process is produced, and each p different characteristic $x_1, x_2, \cdots, x_p$ has a normal distribution. Then the joint probability distribution of the vector $\mathbf{x} = \{x_1, x_2, \cdots, x_p\}$ in control has p-variate normal distribution with known mean vector μ_0 and variance-covariance matrix Σ_0. A sample mean vector $\bar{\mathbf{x}}$ which is subgroup of size n has p-variate normal distribution with known μ_0 and Σ_0/n and the statistic H^2 as given

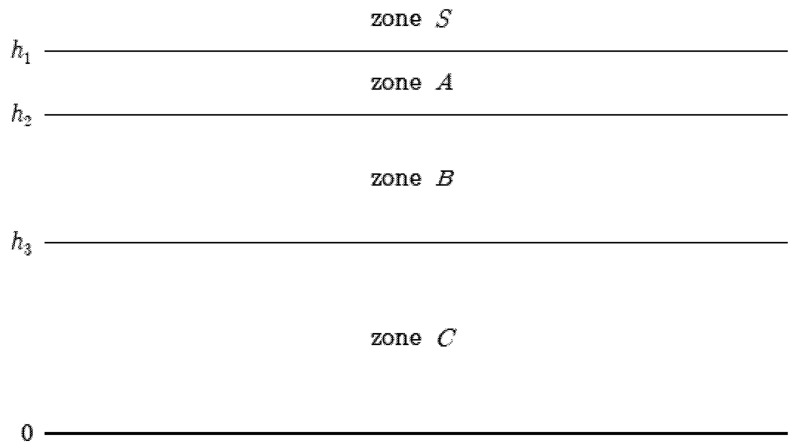

FIGURE 2. Four zones for for multivariate control chart

by

$$H^2 = n(\bar{\mathbf{x}} - \mu_0)'\Sigma_0^{-1}(\bar{\mathbf{x}} - \mu_0) \tag{9}$$

has a chi-square distribution with p degrees of freedom. Suppose that the mean vector in the process changes from μ_0 to μ_1. Set $\mu_1 = \mu_0 + \delta$, then the statistic H^2 has non-central chi-square distribution with p degrees of freedom and non-central parameter(ncp) λ where λ is

$$\lambda = n(\mu_1 - \mu_0)'\Sigma_0^{-1}(\mu_1 - \mu_0) = n\delta'\Sigma_0^{-1}\delta. \tag{10}$$

We will divide the zone into four to apply the rules in Section 3.1 to the chi-square distribution. First zone S has probability which occupied approximately 0.27% of the probability of greater than 3 or less than -3 of the standard normal distribution. Second zone A has the probability which occupied approximately 4.28% of the probability of 2 to 3 and –3 to -2 of the standard normal distribution. Third zone B has the probability which occupied approximately 27.18% of the probability of 1 to 2 and –2 to -1 of the standard normal distribution. Last zone C has the probability which occupied approximately 68.27% of the probability of -1 to -1 of the standard normal distribution. In the chi-square distribution with p degrees of freedom, let h_1 be a value that matches $\chi^2_{p,1-0.9973} = \chi^2_{p,0.0027}$, h_2 be a value that matches $\chi^2_{p,1-0.9545} = \chi^2_{p,0.0455}$, and h_3 be a value that matches $\chi^2_{p,1-0.6827} = \chi^2_{p,0.3173}$. That is, Zone A ranges from h_2 to h_1, Zone B ranges from h_3 to h_2, zone C ranges from 0 to h_3, and zone S is greater than h_1. The zones in the chi-square distribution are shown in Figure 2.

Later, we will combine each rule 2,$\cdots$,6 from section 3.1 with rule 1, in this case applying rule 4 in the same way is not applicable because it has the full range of the chi-square distribution. So, using the notation of Champ and Woodall(1987), we define the multivariate runs rules that apply to the chi-square distribution as follows:

Rule 1 : $MR_1 = T(1, 1, S)$
Rule 2 : $MR_2 = T(2, 3, A)$
Rule 3 : $MR_3 = T(4, 5, A \cup B)$
Rule 4 : $MR_4 = T(2, 2, A)$
Rule 5 : $MR_5 = T(5, 5, A \cup B)$

Using these multivariate runs rules and finite Markov chain embeddings, the run-length distribution for multivariate cohtrol chart was obtained.

3.3. **Distribution of Runs-Length for Multivariate Control Chart with Runs Rules.** We will combine Rule 1 and other rules to obtain a run-length distribution. Combined rules are expressed as MR_{12}, MR_{13}, MR_{14} and MR_{15}. Using these rules, we try to obtain the run-length distribution in a multivariate control chart with a chi-square distribution with p degrees of freedom, *i.e.* p different characteristics. We will only provide the results when the degrees for freedom p is 2, 3 or 7. To apply rules for embedded finite markov chain, we need to re-divide four zones in Section 3.2. We have already assigned a probability for four zones, so we just re-divide the zone. Let Z_1 be the zone where the next test statistic falls to the place specified in each rule, and let Z_2 be between 0 and Z_1. Then, the probability that the next point falls to Z_i is p_i, and the probability of falling to S is p_s. The areas Z_1 and Z_2 are A and $B \cup C$ on the rules MR_{12} and MR_{14}, $A \cup B$ and C on the rules MR_{13} and MR_{15}. You can see these in the Figures 3 and 4. Using the formula of Section 2.2 and the transition probability matrix, we will draw the exact run-length distribution. If a shift occurs in the mean vector with length p, it is defined as $\mu_1 = \mu_0 + \delta$ and has non-central chi-square distribution with p and λ.

So we obtained the run-length distribution for multivariate control charts with degrees of freedom p =2,3 and 7 when this λ changes, and present the

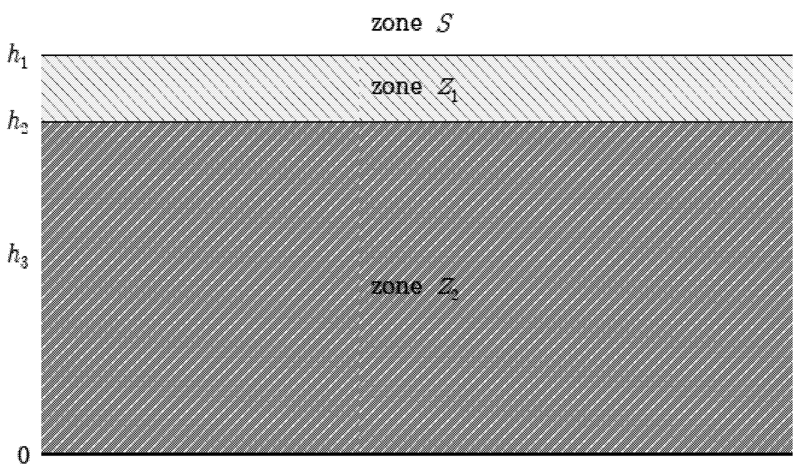

FIGURE 3. Areas of MR_{12} and MR_{14}

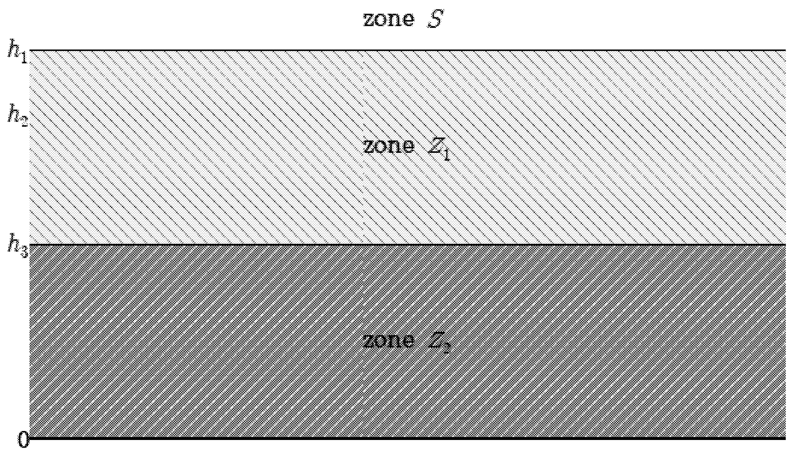

FIGURE 4. Areas of MR_{13} and MR_{15}

ARLs and quartiles following multivariate rules in Table 1. Also, Figure 5 $\sim$ Figure 16 show the probability and cumulative run-length distribution for the multivariate control chart with degrees of freedoms and rules when the λ changes.

Table 1. The values of ARL and quartiles of run-length distribution with Multivariate Rules

		MR_{12}				MR_{13}			
$p(df)$	λ(ncp)	ARL	Q_1	Med	Q_3	ARL	Q_1	Med	Q_3
2	0.0	166.56	49	116	230	53.28	17	38	73
	1.0	29.95	9	21	41	15.42	6	11	20
	5.0	3.76	2	3	5	3.73	2	4	5
	15.0	1.32	1	1	2	1.39	1	1	2
3	0.0	166.56	49	116	230	53.28	17	38	73
	1.0	37.89	12	27	52	17.87	7	13	24
	5.0	4.58	2	3	6	4.21	3	4	5
	15.0	1.42	1	1	2	1.52	1	1	2
7	0.0	166.56	49	116	230	53.28	17	38	73
	1.0	59.16	18	41	82	23.91	9	17	32
	5.0	7.62	3	6	10	5.71	4	5	7
	15.0	1.77	1	2	2	2.00	1	2	3
		MR_{14}				MR_{15}			
$p(df)$	λ(ncp)	ARL	Q_1	Med	Q_3	ARL	Q_1	Med	Q_3
2	0.0	224.39	65	156	311	207.52	61	144	287
	1.0	39.08	12	27	54	37.27	12	26	51
	5.0	4.21	2	3	6	4.74	2	5	6
	15.0	1.32	1	1	2	1.39	1	1	2
3	0.0	224.39	65	156	311	207.52	61	144	287
	1.0	49.87	15	35	69	46.43	15	33	64
	5.0	5.25	2	4	7	5.68	3	5	7
	15.0	1.43	1	1	2	1.54	1	1	2
7	0.0	224.39	65	156	311	207.52	61	144	287
	1.0	79.00	23	55	109	71.22	22	50	98
	5.0	9.20	3	7	12	9.07	5	7	12
	15.0	1.81	1	2	2	2.08	1	2	3

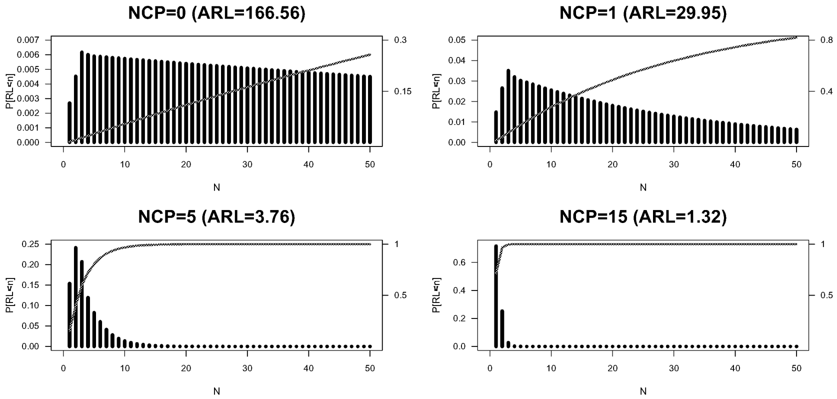

FIGURE 5. Distribution of Run-Length with MR_{12} and p=2

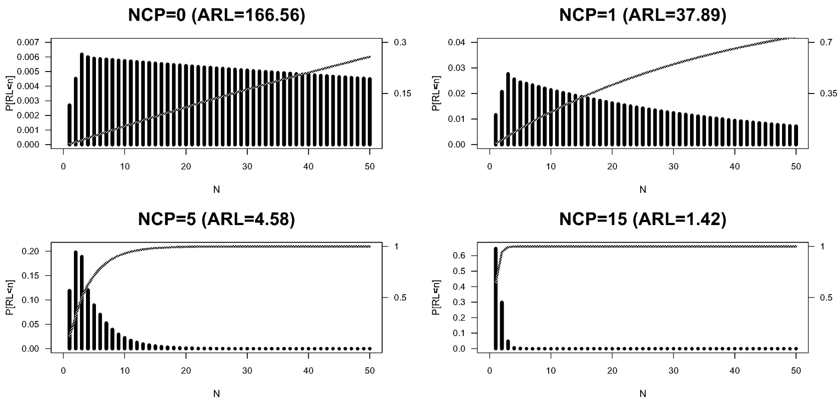

FIGURE 6. Distribution of Run-Length with MR_{12} and p=3

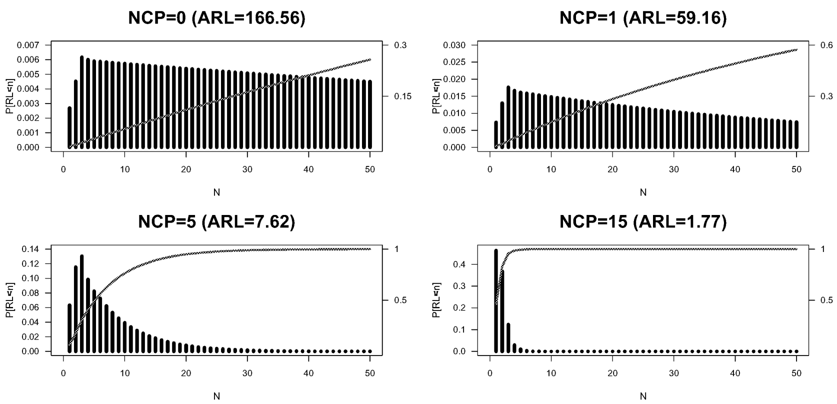

FIGURE 7. Distribution of Run-Length with MR_{12} and p=7

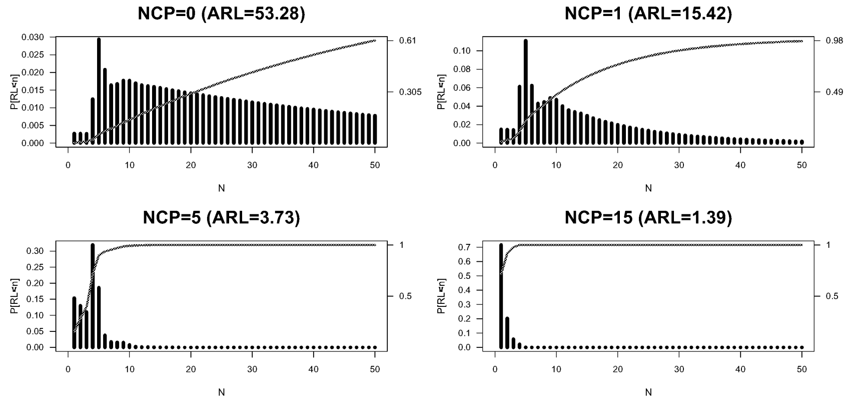

FIGURE 8. Distribution of Run-Length with MR_{13} and p=2

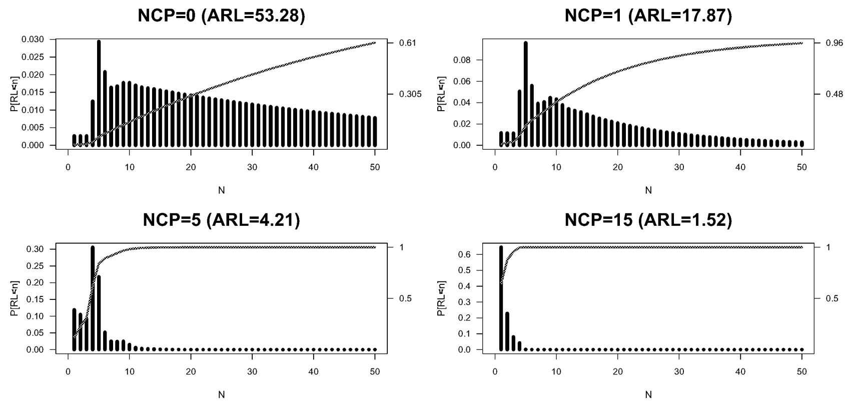

FIGURE 9. Distribution of Run-Length with MR_{13} and p=3

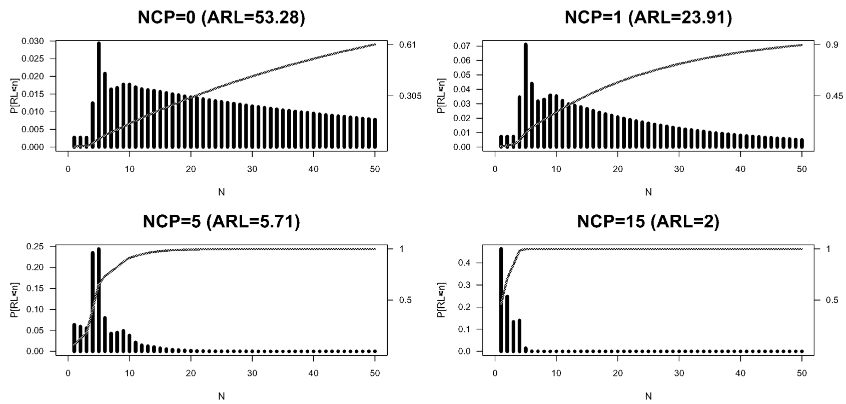

FIGURE 10. Distribution of Run-Length with MR_{13} and p=7

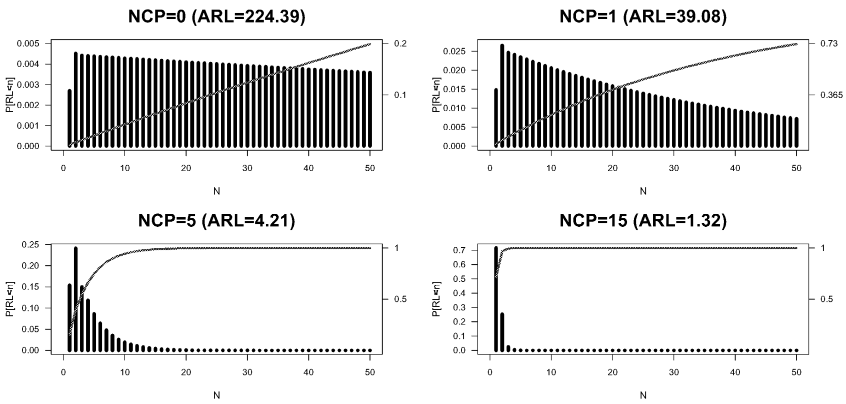

FIGURE 11. Distribution of Run-Length with MR_{14} and p=2

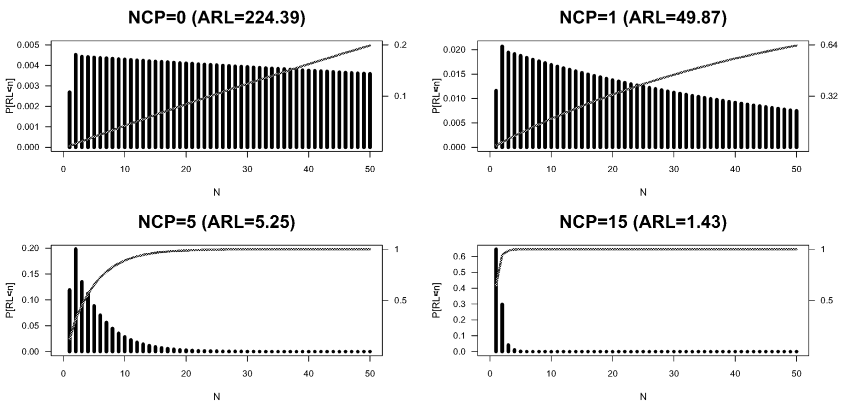

FIGURE 12. Distribution of Run-Length with MR_{14} and p=3

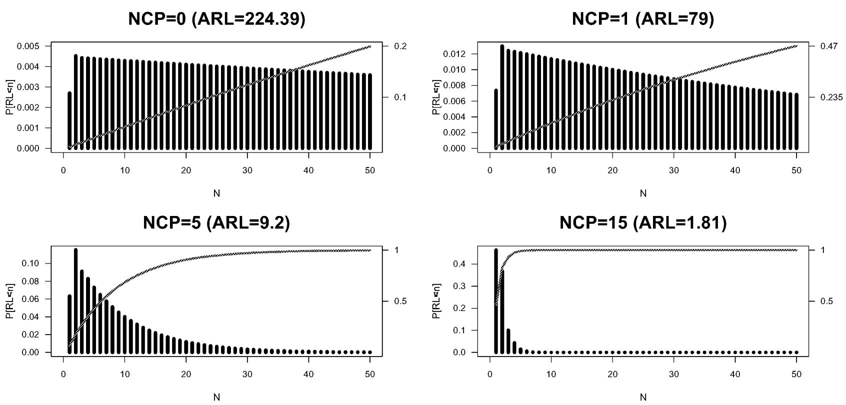

FIGURE 13. Distribution of Run-Length with MR_{14} and p=7

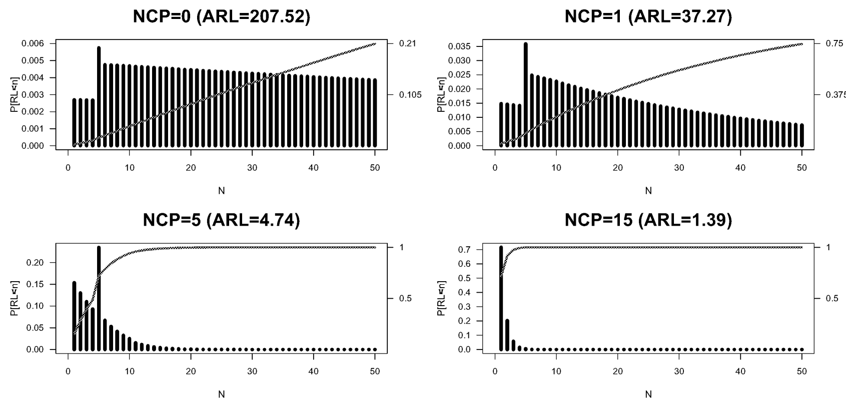

FIGURE 14. Distribution of Run-Length with MR_{15} and p=2

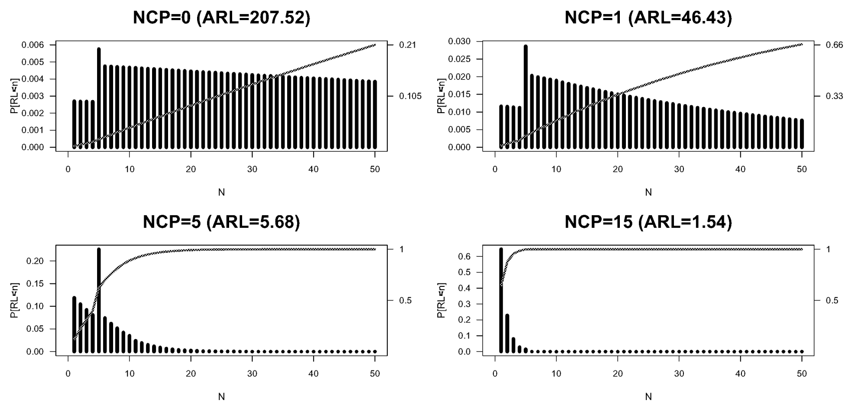

FIGURE 15. Distribution of Run-Length with MR_{15} and p=3

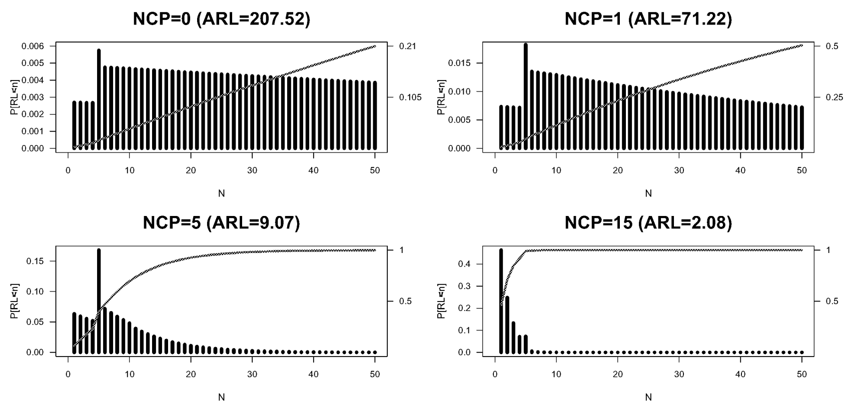

FIGURE 16. Distribution of Run-Length with MR_{15} and p=7

4. Conclusions

We can more easily compute the ARLs and quartiles of run-length distribution by verifying that the ARLs and quartiles of the control charts we calculated using runs rules and finite Markov chain embedding are the same as those obtained by other methods. Run-length distributions of the multivariate control charts obtained in the same way, using runs rules and finite Markov chain embedding, are expected to be accurate, and ARL and quartiles are reduced when the assignable causes occurs in the process. Also, for very small changes, the ARL of the run-length distribution has been shown to decrease rapidly, but it has been found that the ARL is slower as the change increases. Therefore, when we want to control the quality of very sensitive changes, we expect to be able to manage it by using the method proposed so far.

Acknowledgement

Congratulations to Professor Seog-Hoon Rim's retirement ceremony. Thank you for your kind and warm words and great advice. We hope you will always be healthy and happy in the future.

References

[1] Aki, S. (1985). Discrete distributions of order k on a binary sequence, *Ann.Inst. Statist. Math*, **37**, 205-224.
[2] Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*, John Wiley & Sons Inc.
[3] Bissell, A. F. (1978). An Attempt to Unify the Theory of Quality Control Procedures. *Bulletin in Applied Statictics*, **5**, 113-128.
[4] Champ, C. W. and Woodall, W. H. (1987). Exact results for Shewhart control charts with supplementary runs rules, *Technometrics*, **29**, 393-399.
[5] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons Inc.
[6] Fu, J. C. and Koutras, M. V. (1994). Distribution Theory of Runs: A Markov Chain Approach, *Journal of the American Statistical Association*, **89**, 1050-1058.
[7] Fu, J. C. and Lou, W. Y. W. (2003). *Distribution theory of runs and patterns and its applications handbook*, World Scientific Publishing Co.
[8] Hotelling, H. (1947). Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights, *Techniques of Statistical Analysis, McGraw Hill, New York*, 111-184.
[9] Khoo, M. B. C. and Quah, S. H. (2003). Incorporating runs rules into Hotelling's control charts, *Quality Engineering*, **15**, 671-675.
[10] Koutras, M. V. (1996). On a waiting time distribution in a sequence of Bernoulli trials, *Annals of the Institute of Statistical Mathematics*, **48**, 789-806.
[11] Nelson, L. S. (1999). Notes on the Shewhart Control Chart, *Journal of quality technology*, **31**, 124-126.
[12] Page, E. S. (1954). Continuous Inspection Scheme, *Biometrika*, **41**, 100-115.
[13] Philippou, A. N. (1986). Distribution and fibonacci polynomials of order k, longest runs, and reliability of consecutive-k-out-of-n F systems. Fibonacci Numbers and Their Applications, *Reidel Publishing Co*, 203-227.
[14] Roberts, S. W. (1959). Control Chart Tests Based on Geometric Moving Averages, *Technometrics*, **1**, 239-250.
[15] Shewhart, W. A. (1931). *Economic control of Quality of Manufactured Products*, John Wiley & Sons Inc.

[16] Shmueli, G. and Cohen, A. (2003). Run-Length Distribution for Control Charts with Runs and Scans Rules, *Communications in Statistics-Theory and Methods*, **32**, 475-495.

[17] Western Electric Company (1956). *Statistical Quality Control Handbook*, Western Electric Co.

Department of Statistics, Kyungpook National University, 80 Daehakro, Bukgu, Daegu, 41566, Korea

Email address: lhbskans@naver.com

Corresponding Author, Department of Statistics, Kyungpook National University, 80 Daehakro, Bukgu, Daegu, 41566, Korea

Email address: gycho@knu.ac.kr