Advanced Studies in Contemporary Mathematics
30 (2020), No. 2, pp. 231 – 243

www.jangjeon.or.kr
http://dx.doi.org/10.17777/ascm2020.30.2.231

RUN-LENGTH DISTRIBUTION FOR SHEWHART CONTROL CHART WITH RUNS RULES USING FINITE MARKOV CHAIN IMBEDDING

YONGJIN KIM AND GYO-YOUNG CHO

Abstract. Shewhart developed a control chart to statistically control the production process in 1924. The classic Shewhart chart is effective for large shifts and can be easily implemented in the real production process. However, it is not effective for detecting small shifts. To compensate for this, a control chart with the addition of runs rules was proposed by Western Electric Company (1956). And Champ and Woodall (1987) introduced how to calculate the average run-length in runs rules using a method called Markov chain. We have combined runs rules to compensate for the shortcomings of the traditional control chart, and suggest the finite Markov chain imbedding method to get the run-length distribution. The Shewhart control chart with supplementary runs rules can sensitively detect small shifts in means and variances in the production process, and calculate run-length probability distribution accurately and quickly.

2010 Mathematics Subject Classification. 11T23, 20G40, 94B05.

Keywords and phrases. ARL, finite Markov chain imbedding, mean and variance control chart, runs rules, run-length distribution.

1. Introduction

Shewhart developed a control chart to statistically control the production process in 1924. In general, the the control chart consists of the lower control limit and the upper control limit based on the target value, which is the center line. Usually, control limits are set based on 3-sigma at the center line. The production process is said to be in control when all control statistics of the samples obtained from the process are between control limits. If any control statistic is outside the control limit, the production process is considered to be faulty. In this way, it detects fluctuations in the process to detect defects in advance and improve the production capacity of the process. It is also effective for large shifts and can be easily implemented in the real production process. The control chart has been developed through numerous studies and is now used as an important way in the field of quality control. But it is not effective when it comes to small shifts compared to large ones. To compensate for this, Western Electronic Company (1956) has proposed a control chart with runs rules. This method was constantly developed by Bissell (1978) and Nelson (1999). Champ and Woodall (1987) calculated the average run length (ARL) from runs rules using a method called Markov chain. Shmueli and Cohen (2003) proposed a generation function to calculate ARL.

In this paper, we propose the finite Markov chain imbedding method for calculating the run-length distribution of the control chart with runs rules. This obtains the same results as the calculation method of Champ and Woodall (1987) and the calculation method of Shmueli and Cohen (2003). We propose a control chart with supplementary runs rules for monitoring mean and variance. 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 Imbedding

2.1. **Finite Markov Chain Imbedding.** Fu (1985, 1986), Fu and Hu (1987), Chao and Fu (1989, 1991), Fu and Lou (1991) introduced the finite Markov chain imbedding, a method for finding the distribution of random variable $X_n(\Gamma)$. The term "finite Markov chain imbeddable" describes a random variable and was introduced by Fu and Koutras (1994). Let an index set be $\Upsilon_n = \{1, 2, \cdots, n\}$ and let $\Omega = \{b_1, b_2, \cdots, b_m\}$ be a finite state space. For every $x = 0, 1, \cdots, l_n$, if a finite markov chain $\{H_t : t \in \Upsilon_n\}$ and a finite partition $\{Z_x, x = 1, 2, \cdots, l_n\}$ exist in the defined finite state space Ω with initial probability vector $\boldsymbol{\zeta}_0$, then the non-negative integer-valued random variable $X_n(\Gamma)$ is *finite Markov chain imbeddable.* We have

$$P(X_n(\Gamma) = x) = P(H_n \in Z_x | \boldsymbol{\zeta}_0). \tag{1}$$

Fu and Koutras (1994) used the above definition to calculate the probability as follows. Let $\{M_t\}$ be sequence of $m \times m$ transition probability matrices of the imbedded Markov chain $\{H_t\}$ with initial probability distribution $\boldsymbol{\zeta}_0$, for $t = 1, 2, 3, \cdots, n$. If $X_n(\Gamma)$ is finite Markov chain imbeddable,

$$P(X_n(\Gamma) = x) = \boldsymbol{\zeta}_0 (\prod_{t=1}^{n} \boldsymbol{M}_t) \boldsymbol{L}'(Z_x) \ where \ \boldsymbol{L}(Z_x) = \Sigma_{r: b_r \in C_x} \boldsymbol{o}_r, \tag{2}$$

$\boldsymbol{o}_r$ is a $1 \times m$ unit row vector corresponding to state b_r of the state space Ω.

2.2. **Waiting-Time Distribution.** The geometric distribution of order j is often referred to as the Wating-time distribution for Bernoulli trials and was studied by Aki(1985) and Hirano(1986). Let $\Gamma = S \cdots S$ be a j consecutive successes in the pattern of simplification of j, and define the random variable $V(\Gamma)$ as the waiting time for pattern Γ to occur, *i.e.*

$$V(\Gamma) = \inf\{k : X_{k-j+1} = X_{k-j+2} = \cdots = X_k = S\}. \tag{3}$$

For a given pattern length $j \geq 1$ and a sequence of Bernoulli trials, the distribution of $V(\Gamma)$ is given by

$$P(V(\Gamma) = k) = \boldsymbol{\zeta} \boldsymbol{N}^{k-1} (\boldsymbol{I} - \boldsymbol{N}) \boldsymbol{1}', \ k = 1, 2, 3, \cdots, \tag{4}$$

where $\boldsymbol{\zeta} = (1, 0, \ldots, 0)$ is a $1 \times j$ row vector, $\mathbf{1} = (1, 1, \ldots, 1)$ is a $1 \times j$ row vector and $\boldsymbol{N}$ is the $j \times j$ important transition probability submatrix of

$$(5) \quad \boldsymbol{M} = \begin{matrix} 0 \\ 1 \\ \vdots \\ \vdots \\ j-1 \\ \alpha \end{matrix} \left(\begin{array}{cccccc|c} q & p & 0 & \cdots & 0 & 0 \\ q & 0 & p & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ q & 0 & \cdots & \cdots & 0 & p \\ \hline 0 & 0 & \cdots & \cdots & 0 & 1 \end{array} \right)_{(j+1)\times(j+1)} = \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \mathbf{0} & \mathbf{1} \end{array} \right).$$

The expectation of the waiting-time random variable $V(\Gamma)$ is given by

$$(6) \qquad EV(\Gamma) = \sum_{k=1}^{\infty} \boldsymbol{\zeta} \boldsymbol{N}^{k-1} \mathbf{1'} = \boldsymbol{\zeta}(\boldsymbol{I} - \boldsymbol{N})^{-1} \mathbf{1'}.$$

The ARL and average waiting time have the same value. We verify performance with ARL on control charts with supplementary runs rules.

3. Distribution of Run-Length

3.1. **Run.** Balakrishnan and Koutras (2002) introduced how to define and use runs as subgroups of sequential points. In other words, a run is a test that is repeated until a particular result is consistently successful in a test in which a mutually exclusive event occurs. For example, if we have the binary sequence SSSFFFSSF, we can know the following runs. First, we can have a run of the three S 's and a run of three F 's, a run of two S 's run, finally a run of one F. As a result, the total sequence is four.

3.2. **Runs Rules.** To complement the Shewhart control chart, the Western Electric Company (1956) proposed a runs rules as follows

Rule 1: One point is out of the 3 sigma limits
Rule 2: Two of three continuous points are out of the 2 sigma limits
Rule 3: Four of five continuous points are out of the 1 sigma limits
Rule 4: Eight continuous points in one direction relative to the center line

Runs rules apply after dividing the zone in the control chart. Control chart devides into seven parts $(S, A_1, B_1, C_1, C_2, B_2, A_2)$ as Figure 1.

We will use runs and rules notation from Champ and Woodall (1987). They are denoted by $T(k, m, Z)$, which means that if k of the last m standardized points fall in zone Z, it signals. In seven zones, the following shall be considered:

Rule 1: $R_1 = \{T(1, 1, S)\}$
Rule 2: $R_2 = \{T(2, 3, A_2), T(2, 3, A_1)\}$
Rule 3: $R_3 = \{T(4, 5, A_2 \cup B_2), T(4, 5, A_1 \cup B_1)\}$
Rule 4: $R_4 = \{T(8, 8, A_2 \cup B_2 \cup C_2), T(8, 8, A_1 \cup B_1 \cup C_1)\}$
In addition, rules proposed by Duncan(1974):
Rule 5: $R_5 = \{T(2, 2, A_2), T(2, 2, A_1)\}$
Rule 6: $R_6 = \{T(5, 5, A_2 \cup B_2), T(5, 5, A_1 \cup B_1)\}$

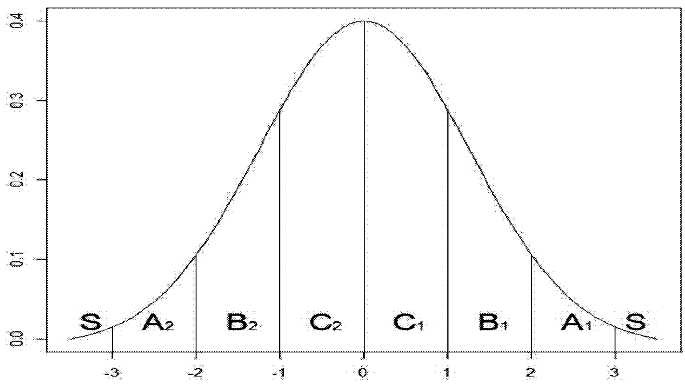

FIGURE 1. 7 zones for control chart

We combined Rule 1 with other rules to create additional runs rules. Previously, runs rules were developed as a Bernoulli trials. In the case of Aki (1985), Hirano (1986) and Koutras (1996), they had studied the waiting-time distribution in the Bernoulli trials according to the geometric distribution of order j.

3.3. **Run-Length Distribution for Mean and Variance Control Chart with Runs Rules.** We previously introduced how ARL can be calculated using the finite Markov chain imbedding.Traditional control chart has not been sensitive to small shifts in the production process. So, we have added runs rules to the control chart to sensitively detect small shifts in the production process. We used finite Markov chain imbedding to calculate the ARL for control charts with runs rules in the production process. We observe the ARLs and quartiles when the mean and variance change due to small shifts.

Assume that the production process has the characteristics of a normal distribution, where mean μ_0 and variance σ_0^2 are known. When the production process is controlled, the control statistic Y_i is as follows:

$$Y_i = \sum_{j=1}^{n} \left(\frac{X_{ij} - \mu_0}{\sigma_0} \right)^2 = \frac{(n-1)S_i^2}{\sigma_0^2} + \frac{n(\bar{x}_i - \mu_0)^2}{\sigma_0^2} \sim \chi_n^2, \tag{7}$$

and it has a chi-square distribution with n degrees of freedom.

Suppose that the mean and variance in the process change from μ_0 to μ_1 and σ_0^2 to σ_1^2, then the control statistic Y_i is as

$$\begin{aligned} Y_i = \sum_{j=1}^{n} \left(\frac{X_{ij} - \mu_0}{\sigma_0} \right)^2 &= \sum_{j=1}^{n} \left(\frac{(X_{ij} - \mu_1) + (\mu_1 - \mu_0)}{\sigma_1} \frac{\sigma_1}{\sigma_0} \right)^2 \\ &= \frac{\sigma_1^2}{\sigma_0^2} \sum_{j=1}^{n} \left(\frac{X_{ij} - \mu_1}{\sigma_1} + \frac{\mu_1 - \mu_0}{\sigma_1} \right)^2 \\ &\sim \frac{\sigma_1^2}{\sigma_0^2} \chi^2_{(n,\lambda)}, \quad \lambda = n\frac{(\mu_1 - \mu_0)^2}{\sigma_1^2} \end{aligned} \tag{8}$$

and it has a non-central chi-square distribution with n degrees of freedom and the non-centrality parameter λ.

Through this, we obtain the ARLs and quartiles when means and variances change.

We will divide the zone into four to apply the runs rules to the chi-square distribution. Each zone is S, A, B, C as following and the probability is 0.0027, 0.0428, 0.2718 and 0.6827 respectively.

We use the notation of Camp and Woodall (1987) to use the following runs rules:

Rule 1: $R_1 = \{T(1, 1, S)\}$
Rule 2: $R_2 = \{T(2, 3, A)\}$
Rule 3: $R_3 = \{T(4, 5, A \cup B)\}$
In addition, rules proposed by Duncan(1974):
Rule 4: $R_4 = \{T(2, 2, A)\}$
Rule 5: $R_5 = \{T(5, 5, A \cup B)\}$

We combine Rule 2, Rule 3, Rule 4, and Rule 5 based on Rule 1 to create new runs rules.

Let p_i be the probability of entering each zone. If we combine Rule 1 with Rule 2, then the possible states are S, BS_cS_c, S_cBS_c, CS_cS_c, S_cCS_c, S_cS_c where $S_c = A$ in Figure 2. The probability of transition matrix M is $p_s = \delta = 0.0027$, $p_{s_c} = P(S_c) = P(A) = 0.0428$ and $p_b = P(B) = 0.2718$ and $p_c = P(C) = 0.6827$.

If we combine Rule 1 with Rule 3, then the possible states are S, $CS_cS_cS_cS_c$, $S_cCS_cS_cS_c$, $S_cS_cCS_cS_c$, $S_cS_cS_cCS_c$, $S_cS_cS_cS_c$ where $S_c = A \cup B$ in Figure 2. The probability of transition matrix M is $p_s = \delta = 0.0027$, $p_{s_c} = P(S_c) = 0.3146$ and $p_c = P(C) = 0.6827$.

If we combine Rule 1 with Rule 4, then the possible states are S, S_cS_c where $S_c = A$ in Figure 2. The probability of transition matrix M is $p_s = \delta = 0.0027$, $p_{s_c} = P(S_c) = P(A) = 0.0428$ and $p_b = P(B) = 0.2718$ and $p_c = P(C) = 0.6827$.

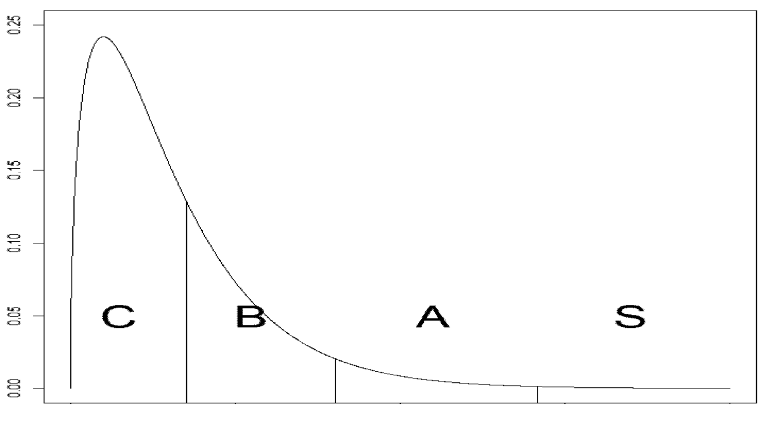

FIGURE 2. 4 zones of control chart for chi-square distribution

If we combine Rule 1 with Rule 5, then the possible states are S, $S_cS_cS_cS_cS_c$ where $S_c = A \cup B$ in Figure 2. The probability of transition matrix M is $p_s = \delta = 0.0027$, $p_{s_c} = P(S_c) = 0.3146$ and $p_c = P(C) = 0.6827$.

We show the ARLs and quarties when the means and variances change. If the mean and varince change from μ_0 to μ_1 and σ_0^2 to σ_1^2 respectively, then the distribution of Y_i is the non-central chi-square with $\sigma_1^2/\sigma_0^2\chi^2_{(n,\lambda)}$, $\lambda = n(\mu_1 - \mu_0)^2/\sigma_1^2$.

In this paper, r is σ_1/σ_0, λ is $n(\mu_1 - \mu_0)^2/\sigma_1^2$. We set conditions that degrees of freedom are 3,5, r=1,1.5 and λ=0.0, 2.0, 4.0. Table 1 $\sim$ Table 4 show ARLs and the quartiles of run-length of the control chart with runs rules when the mean and varince change from μ_0 to μ_1 and σ_0^2 to σ_1^2, respectively.

Table 1. The values of ARL and quartiles of run-length distribution with Rule 1 and Rule 2

df=3						df=5					
r	λ	ARL	Q_1	Med	Q_3	r	λ	ARL	Q_1	Med	Q_3
	0.0	166.60	49	116	230		0.0	166.58	49	116	230
1	2.0	16.41	6	12	22	1	2.0	22.91	7	16	31
	4.0	6.23	3	5	8		4.0	8.53	3	6	11
	0.0	6.45	3	5	9		0.0	4.39	2	3	6
1.5	2.0	2.70	1	2	3	1.5	2.0	2.41	1	2	3
	4.0	1.83	1	2	2		4.0	1.76	1	1	2

Table 2. The values of ARL and quartiles of run-length distribution with Rule 1 and Rule 3

df=3						df=5					
r	λ	ARL	Q_1	Med	Q_3	r	λ	ARL	Q_1	Med	Q_3
	0.0	53.29	17	38	73		0.0	53.28	17	38	73
1	2.0	9.75	5	7	13	1	2.0	12.06	5	9	16
	4.0	5.12	4	5	6		4.0	6.19	4	5	8
	0.0	5.66	3	5	7		0.0	4.19	2	4	5
1.5	2.0	2.86	1	3	4	1.5	2.0	2.59	1	2	4
	4.0	1.97	1	2	3		4.0	1.90	1	1	2

TABLE 3. The values of ARL and quartiles of run-length distribution with Rule 1 and Rule 4

df=3						df=5					
r	λ	ARL	Q_1	Med	Q_3	r	λ	ARL	Q_1	Med	Q_3
1	0.0	224.46	65	156	311	1	0.0	224.42	65	156	311
	2.0	20.91	7	15	29		2.0	29.72	9	21	41
	4.0	7.37	3	5	10		4.0	10.41	4	7	14
1.5	0.0	7.36	3	5	10	1.5	0.0	4.87	2	4	6
	2.0	2.84	1	2	4		2.0	2.52	1	2	3
	4.0	1.87	1	2	2		4.0	1.79	1	1	2

TABLE 4. The values of ARL and quartiles of run-length distribution with Rule 1 and Rule 5

df=3						df=5					
r	λ	ARL	Q_1	Med	Q_3	r	λ	ARL	Q_1	Med	Q_3
1	0.0	207.58	61	145	287	1	0.0	207.54	61	145	287
	2.0	19.77	7	14	27		2.0	27.15	9	20	37
	4.0	7.61	4	6	10		4.0	10.18	5	8	13
1.5	0.0	7.97	3	6	11	1.5	0.0	5.36	2	5	7
	2.0	3.17	1	3	5		2.0	2.82	1	2	4
	4.0	2.04	1	2	3		4.0	1.96	1	1	2

Figure 3 $\sim$ Figure 10 show the probability and cumulative run-length distribution of the control chart with runs rules when the mean and varince change from μ_0 to μ_1 and σ_0^2 to σ_1^2, respectively.

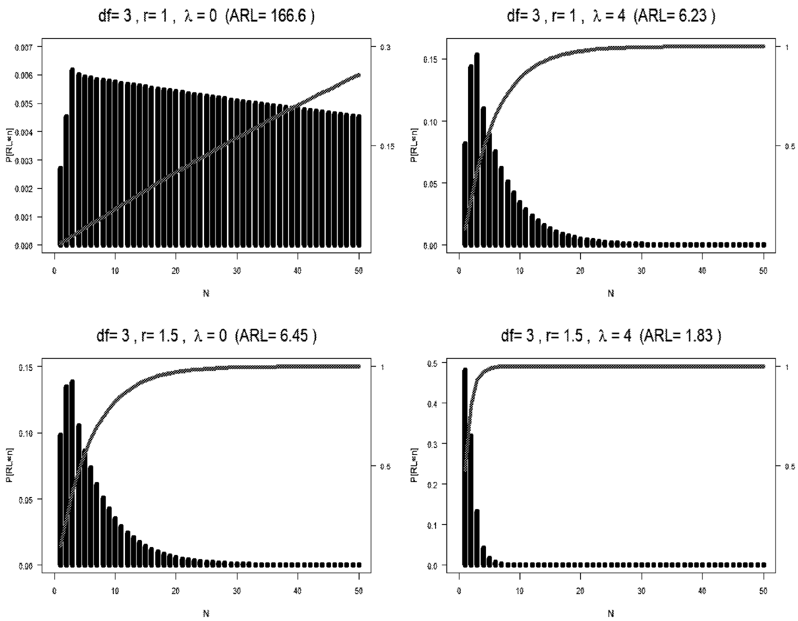

FIGURE 3. Run-length distribution with Rule 1 and Rule 2 (df=3)

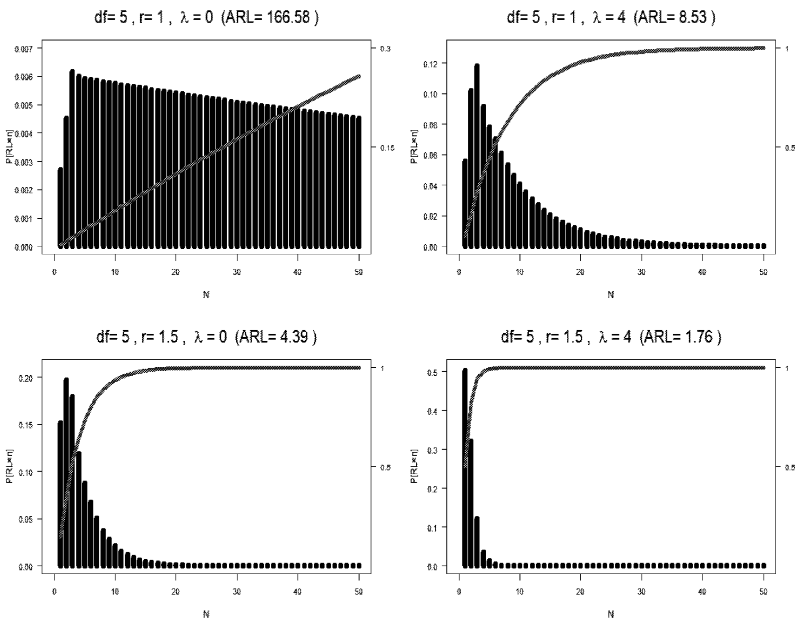

FIGURE 4. Run-length distribution with Rule 1 and Rule 2 (df=5)

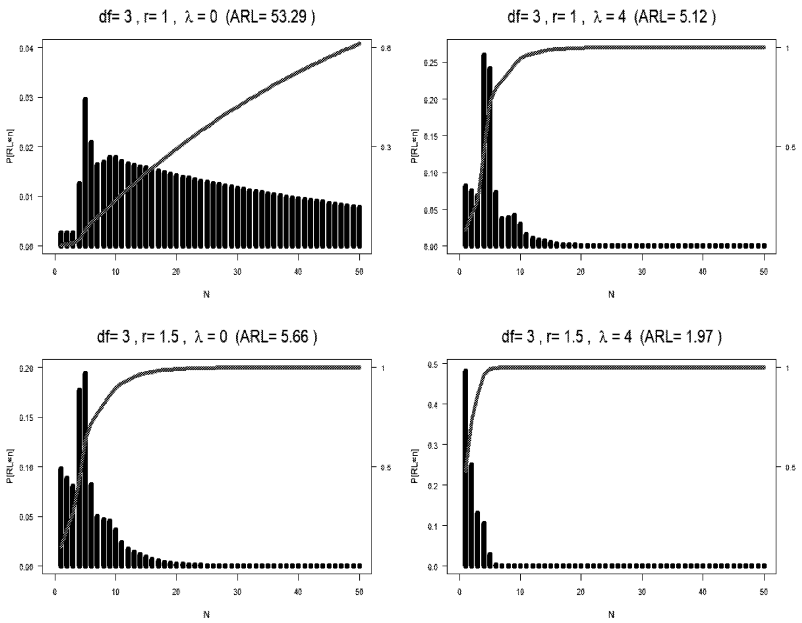

FIGURE 5. Run-length distribution with Rule 1 and Rule 3 (df=3)

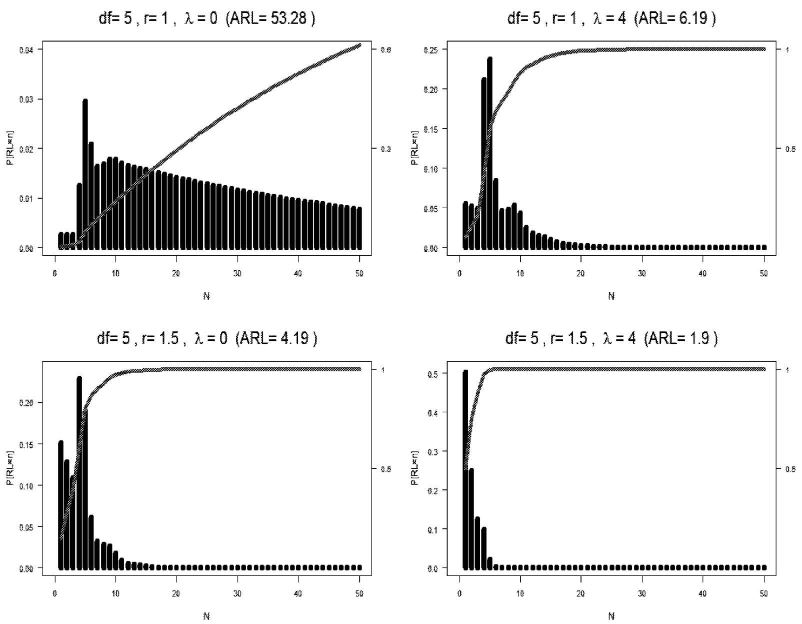

FIGURE 6. Run-length distribution with Rule 1 and Rule 3 (df=5)

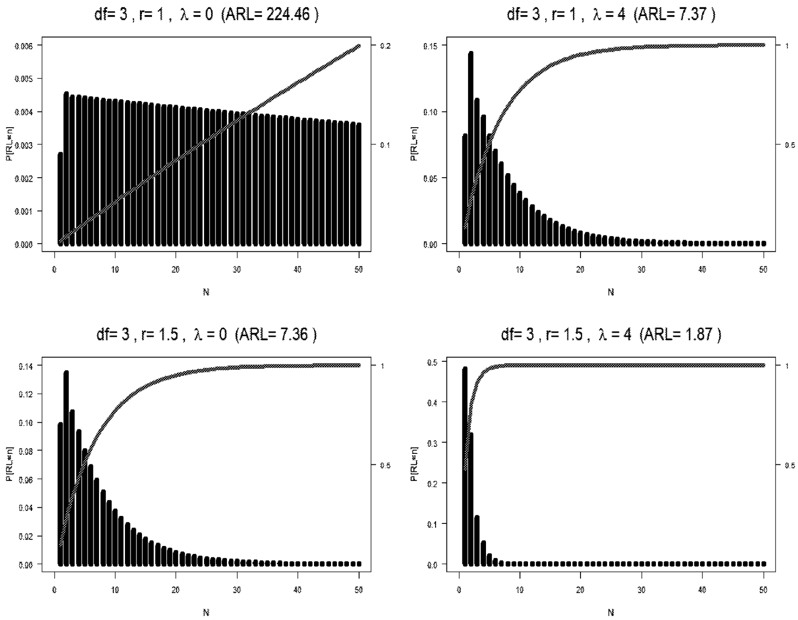

FIGURE 7. Run-length distribution with Rule 1 and Rule 4 (df=3)

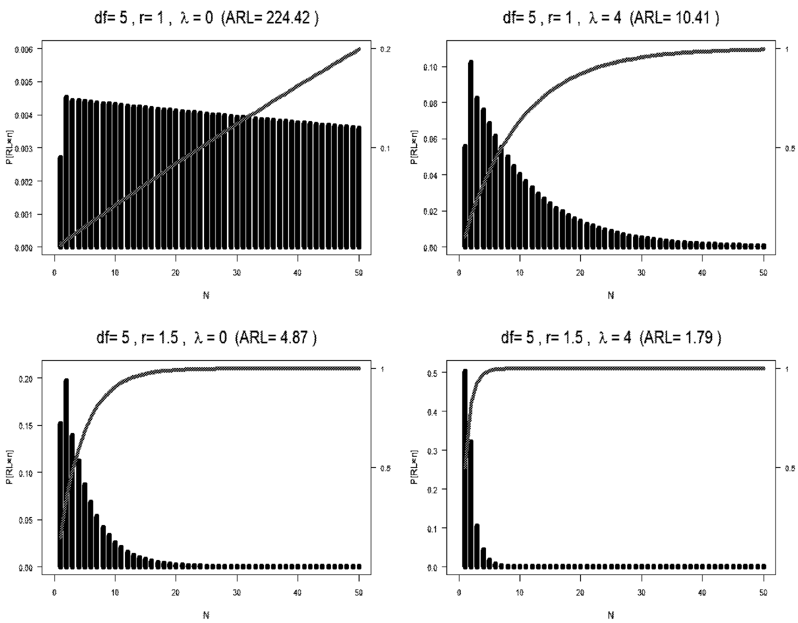

FIGURE 8. Run-length distribution with Rule 1 and Rule 4 (df=5)

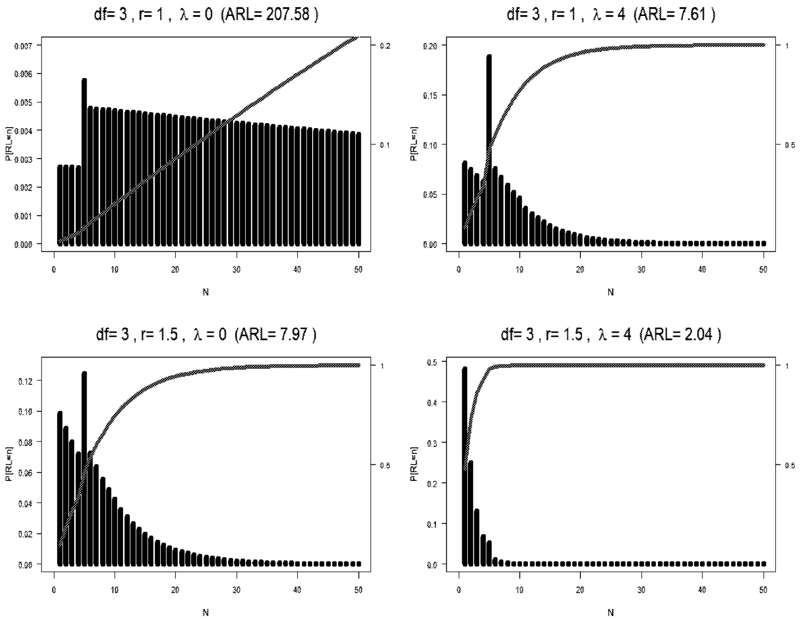

Figure 9. Run-length distribution with Rule 1 and Rule 5 (df=3)

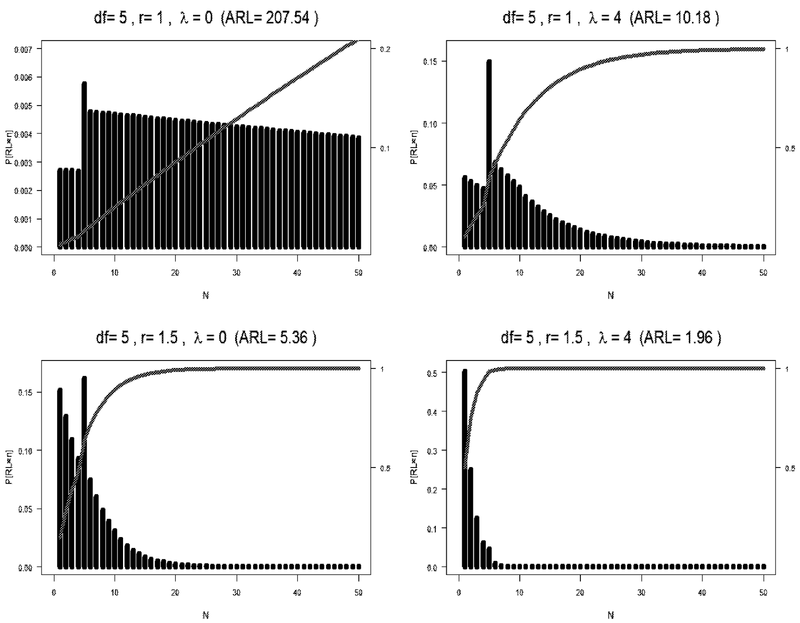

Figure 10. Run-length distribution with Rule 1 and Rule 5 (df=5)

4. Conclusions

We have combined runs rules to compensate for the shortcomings of the traditional control chart, and suggest the finite Markov chain imbedding method to get the run-length distribution. The Shewhart control chart with supplementary runs rules can sensitively detect small shifts in means and variances in the production process, and we can calculate run-length probability distribution accurately and quickly. The Shewhart control chart with supplementary runs rules is effective in detecting small shifts in means and variances.

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] Chao, M. T. and Fu, J. C. (1989). A limit theorem of certain repairable systems, *Annals of the Institute of Statistical Mathematics*, **41**, 809-818.

[6] Chao, M. T. and Fu, J. C. (1991). The reliability of large series system under a Markovian structure, *Advances in Applied Probability*, **23**, 894-908.

[7] Duncan, A. J. (1974). *Quality control and industrial statistics*, R. D. Irwin; 4 edition.

[8] Fu, J. C. (1985). Reliability of consecutive-k-out-of-n:F system, *IEEE Transactions on Reliability*, **34**, 127-130.

[9] Fu, J. C. (1986). Reliability of consecutive-k-out-of-n:F system sith $(k-1)$ step Markov dependence, *IEEE Transactions on Reliability*, **35**, 602-606.

[10] Fu, J. C. and Hu, B. (1987). On reliability of a large consecutive-k-out-of-n:F system with $k|1$ step Markov dependenc, *IEEE Transactions on Reliabilit*, **36**, 75-77.

[11] 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.

[12] Fu, J. C. and Lou, W. Y. W. (1991). On reliabilities of certain large linearly connected engineering systems, *Statistics and Probability Letters* , **12**, 291-296.

[13] Hirano, K. (1986). Some properties of the distributions of order k, Fibonacci numbers and their applications (eds. A.N. Philippou, G.E.Bergum and A.F. Horadam), *Reidel, Dordrecht*, 43-53.

[14] 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.

[15] Nelson, L. S. (1999). Notes on the Shewhart Control Chart, *Journal of quality technology*, **31**, 124-126.

[16] Shewhart, W. A. (1931). *Economic control of Quality of Manufactured Products*, John Wiley & Sons Inc.

[17] 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.

[18] 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: hoozama12@naver.com

Corresponding Author, Department of Statistics, Kyungpook National University, 80 Daehakro, Bukgu, Daegu, 41566, Korea
Email address: gycho@knu.ac.kr