Chi-square Control Charts with Runs Rules

Article *in* Methodology and Computing in Applied Probability · December 2011

DOI: 10.1007/s11009-010-9178-7

2 authors:

Athanasios C. Rakitzis
University of the Aegean

Demetrios Antzoulakos
University of Piraeus

Methodol Comput Appl Probab (2011) 13:657–669
DOI 10.1007/s11009-010-9178-7

Chi-square Control Charts with Runs Rules

Athanasios C. Rakitzis · Demetrios L. Antzoulakos

Received: 14 November 2009 / Revised: 5 February 2010
Accepted: 6 April 2010 / Published online: 8 May 2010

Abstract The Hotelling's χ^2 control chart is one of the most widely used multivariate charting procedures for monitoring the vector of means of several quality characteristics. As a Shewhart-type control chart, it incorporates information pertaining to most recently inspected sample and subsequently it is relatively insensitive in quickly detecting small magnitude shifts in the process mean vector. A popular solution suggested to overcome this handicap was the use of runs and scans rules as criteria to declare a process out-of-control. During the last years, the examination of Hotelling's χ^2 control charts supplemented with various runs rules has attracted continuous research interest. In the present article we study the performance of the Hotelling's χ^2 control chart supplemented with a r-out-of-m runs rule. The new control chart demonstrates an improved performance over other competitive runs rules based control charts.

Keywords Average run length · Control charts · Markov chain imbedding technique · Runs rules · Waiting time distribution

AMS (2000) Classification Primary 62P30 · Secondary 62E15

1 Introduction

There are many situations in practice in which it is necessary the simultaneous control of two or more related quality characteristics. Such process monitoring or control problems are referred in the literature as multivariate quality control problems. The work in the area of multivariate quality control was initiated by Hotelling (1947). Since then many papers

A. C. Rakitzis · D. L. Antzoulakos (✉)
Department of Statistics and Insurance Science, University of Piraeus, Karaoli & Dimitriou 80, 18534 Piraeus, Greece
e-mail: dantz@unipi.gr

A. C. Rakitzis
e-mail: arakitz@gmail.com

dealing with control procedures for several related quality characteristics appeared in the literature. Among them we mention Alt and Smith (1988); Crosier (1988); Pignatiello and Runger (1990); Lowry et al. (1992), and references therein. We refer to Bersimis et al. (2007) for a recent review in multivariate control charting procedures.

Hotelling's χ^2 (chi-square) control chart (abbr. CSCC) is the most widely used multivariate control chart for monitoring the mean vector of a process. It is a direct analog of the univariate Shewhart $\overline{X}$ control chart and it gives an out-of-control signal as soon as a point exceeds the upper control limit of the chart. No matter how easy it is in implementation and interpretation, the CSCC is not very sensitive in the detection of small and moderate shifts in the mean vector, since it is based on the most recent observation. A common approach to increase the sensitivity of any Shewhart-type control chart and preserve its simplicity is the use of supplementary rules based on runs and scans which make use of additional information from the recent history of the process (for an up-to-date review on this subject we refer to Koutras et al. 2007).

Khoo and Quah (2003) were the first who incorporate runs rules in the CSCC. They proposed the use of runs rules of the type "*a* out of *b* consecutive points fall in a certain interval" (abbr. a / b runs rule), as a criterion providing evidence that the mean vector has changed. They conducted a simulation study to evaluate the average run length (ARL) performance of the 1 / 1, 2 / 2, 2 / 3 and 2 / 4 runs rules. Prior to incorporating runs rules in the control chart, the plotted statistic was transformed from a chi-square random variable to a standard normal one. In the same spirit, Khoo and Quah (2004) applied Khoo and Quah's (2003) ideas for monitoring process dispersion based on the sample generalized variance. Khoo et al. (2005) extended Khoo and Quah's (2003) approach by combining two a / b runs rules in a control chart with two control limits. Khoo (2005), modified the work of Khoo and Quah (2003) and Khoo et al. (2005) in order to be applicable in the case where the original chi-square statistic is plotted on the control chart instead of a transformed standard normal random variable. Aparisi et al. (2004) investigated the ARL performance of a CSCC supplemented with four runs rules. Finally, Koutras et al. (2006) studied in detail a CSCC supplemented with a m / m runs rule which has better ARL performance than the standard CSCC. In addition, they studied the performance of a CSCC supplemented with a 1 / 1 and a m / m runs rule.

In the present work we propose a modification of the standard CSCC. The new chart gives an out-of-control signal when a single point exceeds a suitable upper outer control limit, or when r points are plotted between the upper outer control limit and a suitable upper inner control limit which are separated by at most $m–r$ points ($2 \leq r < m$) located between the center line and the upper inner control limit of the chart. The new control chart increases CSCC's sensitivity in the detection of small to moderate shifts in the mean vector and allows quicker detection of large ones. The present paper is organized as follows: In Section 2 we introduce the basic features of the new CSCC and present the results of a systematic numerical study regarding its ARL performance. Design aspects of the new CSCC are discussed in Section 3 while conclusions are summarized in Section 4. In the Appendix, we present a Markov chain approach suitable for the study of the run length distribution of runs rules based control charts.

2 The CS: *r* / *m* Control Chart

Consider a process in which p correlated quality characteristics, $x_1,$ x_2, ... x_p, are being monitored simultaneously. Assume that the in-control joint probability distribution of the vector $\mathbf{x} = (x_1,x_2,...,x_p)$ follows the p-variate Normal distribution with known in-control mean vector $\boldsymbol{\mu}_0$ and variance-covariance matrix $\boldsymbol{\Sigma}_0$, that is $\mathbf{x} \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$). Rational subgroups of size $n>1$ are collected sequentially and the mean sample vector $\overline{\mathbf{x}}_i$ of the i-th subgroup is evaluated. In a CSCC the subgroup statistics

$$T_i^2 = n(\overline{\mathbf{x}}_i - \boldsymbol{\mu}_0) \sum\nolimits_0^{-1} (\overline{\mathbf{x}}_i - \boldsymbol{\mu}_0), \quad i \geq 1,$$

are plotted on the chart in a sequential order. In case of individual observations ($n=1$), $\overline{\mathbf{x}}_i$ should be replaced by $\mathbf{x}_i$, $i \geq 1$.

For an in-control-process, the plotted statistic follows a chi-square distribution with p degrees of freedom, that is $T_i^2 \sim \chi_p^2$, $i \geq 1$. Consequently, for a probability of a false alarm rate of a, the upper control limit (UCL) of the CSCC is the upper a-percentage point of the chi-square distribution with p degrees of freedom, that is $\mathrm{UCL} = \chi_{p,a}^2$. The CSCC gives an out-of-control signal when a single plotted point exceeds UCL. We will refer to this classical rule of obtaining an out-of-control signal as the 1 / 1 (runs) rule.

Assume that the appearance of an assignable cause affects only the mean vector of the process by producing a shift in at least one of its components, while the variance-covariance matrix $\boldsymbol{\Sigma}_0$ remains on a stable and unchanged state. When the in-control mean vector $\boldsymbol{\mu}_0$ shifts to $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \boldsymbol{\delta}$ ($\boldsymbol{\delta} \neq \mathbf{0}$), the magnitude of this shift is often expressed by the Mahalanobis distance

$$d = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \sum\nolimits_0^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)'} = \sqrt{\boldsymbol{\delta} \sum\nolimits_0^{-1} \boldsymbol{\delta}'}.$$

The subgroup statistic T_i^2 then follows a non-central chi-square distribution with p degrees of freedom and non-centrality parameter $\lambda = nd^2$, that is $T_i^2 \sim \chi_p^2(\lambda)$.

A commonly used performance indicator for a control chart is its ARL. It is the average number of points plotted on a control chart until an out-of-control signal is obtained. In the case of the CSCC the ARL is given by

$$\mathrm{ARL} = \frac{1}{\Pr(T^2 > \mathrm{UCL}|d)}.$$

It is well-known that the CSCC is not very sensitive in the detection of small and moderate shifts in the mean vector. To overcome this weakness, maintaining at the same time its standard performance in detecting large shifts, we suggest the use of two runs rules as a criterion for obtaining an out-of-control signal.

Consider a CSCC with a center line CL, an upper inner control limit (UICL) and an upper outer control limit (UOCL) satisfying the inequality $0 < \mathrm{CL} < \mathrm{UICL} < \mathrm{UOCL}$. The control chart gives an out-of-control signal if either a single point exceeds UOCL, or r points are plotted between the UICL and UOCL which are separated by at most m–r points located between CL and UICL ($2 \leq r < m$). The center line of the proposed chart is defined to be the median of the in-control distribution of the plotted statistic, that is

$$\Pr\left(T^2 > \mathrm{CL}|d = 0\right) = 0.5.$$

The aforementioned criterion providing evidence that the process mean vector is out-of-control is a combination of the 1 / 1 rule used in the standard CSCC and a r / m runs rule. We will refer to this new control chart as the "*r-out-of-m chi-square control chart*", to be denoted by CS: r / m.

A graphical representation of the CS: 2 / 4 control chart is given in Fig. 1. We observe that point 20 gives an out-of-control signal since points 17 and 20 are located between

UOCL and UICL, and points 18 and 19 are located between CL and UICL. Furthermore, point 10 gives an out-of-control signal since it is located above UOCL, while point 5 does not give an out-of-control signal since point 3 is not located between CL and UICL.

In a CS: r / m control chart we may distinguish four different regions: one extending below CL (region 0), one extending between CL and UICL (region 1), one extending between UICL and UOCL (region 2), and one extending above UOCL (region 3). The probabilities p_0, p_1, p_2 and p_3 that a single plotted point falls in regions 0, 1, 2 and 3 are given by

$$\begin{aligned} p_0 &= \Pr(T^2 \leq \mathrm{CL}|d),\ p_1 = \Pr(\mathrm{CL} < T^2 \leq \mathrm{UICL}|d), \\ p_2 &= \Pr(\mathrm{UICL} < T^2 \leq \mathrm{UOCL}|d),\ p_3 = \Pr(T^2 > \mathrm{UOCL}|d). \end{aligned} \tag{4.1}$$

Following a Markov chain approach we may derive the ARL of the CS: r / m control chart (see the Appendix for further details). For the determination of the two limits of the chart we propose fixing the UOCL and then compute the UICL so as to achieve a predetermined in-control ARL value c, that is $\mathrm{ARL}_0 = c$. A natural choice for the UOCL is the UCL of the standard CSCC, that is $\mathrm{UOCL} = \chi^2_{p,\,a}$ (a<1/c). Consequently, ARL_0 will be a function of the unknown probability p_1 since for an in-control process (d=0) we have that p_0=1/2, $p_3 = a$ and $p_2 = (1/2) - a - p_1$.

An algorithmic description for the design of the CS: r / m control chart in discrete steps is as follows:

Step 1: Choose two positive integers r, m ($2 \leq r < m$), the desired in-control ARL value c and a real number a such that $0<a<0.5$ and $a<1/c$.
Step 2: Set $\mathrm{UOCL} = \chi^2_{p,\,a}$.
Step 3: Calculate the unique root p_1 of the equation $\mathrm{ARL}_0 = c$ in the interval (0, 0.5−a).
Step 4: Set $\mathrm{UICL} = \chi^2_{p,\,0.5-p_1}$.
Step 5: Declare the process out-of-control if either a single point exceeds UOCL, or r points are plotted between the UICL and UOCL which are separated by at most m–r points located between CL and UICL.

In Tables 2 and 3 we present the results of a systematic numerical study regarding the ARL performance of the CS: r / m control chart for $2 \leq r < m \leq 5$. We have selected ARL_0= 200, sample size n=1, number of quality characteristics p=5 and 10, a=300^{-1}, 500^{-1} and 1000^{-1}, and d=0(0.25)3. The results are also valid for other sample sizes taking into account equivalent combinations between n and d giving the same value for the non-centrality parameter λ (see, e.g., Table 1 of Aparisi et al. 2004). The UOCL values of the CS: r / m control chart for the six combinations of p and a are given in the following table.

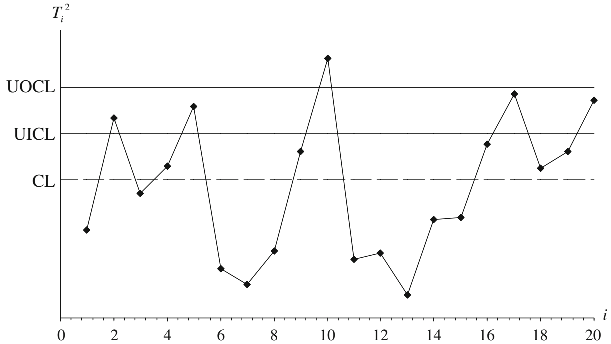

Fig. 1 Graphical representation of the CS: 2 / 4 control chart

For each d value, the column labeled CS: r / m provides the lowest ARL value among all the CS: r / m control charts under investigation. In parenthesis, below the ARL value, we give the characteristics of the CS: r / m control chart succeeding the lowest ARL value in the form (r; m; UICL; UOCL).

For comparison purposes, in Tables 2 and 3 we also give ARLs of the 1 / 1−m / m CSCC studied by Koutras et al. (2006) (column labeled "1 / 1−m / m") and the combined r-of-m and 1-of-1 rule proposed by Khoo et al. (2005) (column labeled "K: r-of-m"). Both control charting procedures utilize a UOCL and a UICL. The 1 / 1−m / m CSCC gives an out-of-control signal if either a point exceeds the UOCL, or m consecutive points are plotted between UICL and UOCL. The K: r-of-m CSCC signals if either a point exceeds the UOCL, or r out of m points are plotted between UICL and UOCL. For the 1 / 1−m / m CSCC we examined the cases m=2, 3, 4 and 5, while for the K: r-of-m CSCC we examined the cases $2 \leq r < m \leq 5$. The UOCL values of both charts were selected from the UOCL values given in Table 1, while the UICL values of the control charts were computed in order to achieve an in-control ARL equal to 200. For each d value we provide the lowest ARL value among all the control charts under investigation. In parenthesis, besides the ARL value, we give the characteristics of the control chart succeeding the lowest ARL value in the form (m; UICL; UOCL) for the 1 / 1−m / m CSCC and in the form (r; m; UICL; UOCL) for the K: r-of-m control chart. Furthermore, the standard CSCC corresponds to the column labeled "1 / 1", while the column labeled "SRR" corresponds to the CSCC supplemented with four runs rules studied by Aparisi et al. (2004). The ARL values in the "SRR" column were evaluated via simulation (see Aparisi et al. (2004) for more details). For each d value, the boldfaced entries in the tables indicate the lowest out-of-control ARL value. Additional tables are available from the authors upon request.

Tables 2 and 3 reveal that CS: r / m control chart has better ARL performance than that of the standard CSCC, the CSCC studied by Aparisi et al. (2004), as well as to the 1 / 1−m / m CSCC introduced by Koutras et al. (2006). Also, our extensive numerical experimentation revealed that for a small number of quality characteristics p and small d values ($d \leq 1$) the K: r-of-m CSCC has better ARL performance than that of the CS: r / m control chart. However as the number of quality characteristics increases the CS: r / m control chart outperforms the K: r-of-m CSCC. Similar conclusions are also valid for various choices of n and ARL_0. Therefore, the CS: r / m control chart can be considered as a viable alternative to the standard CSCC, as well as to the control charts suggested by Aparisi et al. (2004), Khoo et al. (2005) and Koutras et al. (2006).

3 Optimal Control Limits for the CS: *r* / *m* Control Chart

In the previous section we described a procedure for the determination of the two control limits of the CS: r / m control chart by fixing first the UOCL and calculating next the UICL in such a way that ARL_0 possesses the desired value. If we are not willing to pre-specify the

Table 1 Values of UOCL for the CS: r / m control chart (ARL_0= 200)

p	a: 300^{-1}	500^{-1}	1000^{-1}
5	17.710	18.907	20.515
10	26.320	27.722	29.588

Table 2 ARL profiles for ARL_0=200 and p=5

d	1/1	1/1−m/m	K: r-of-m	CS: r / m	SRR
0.00	200	200	200	200	200
0.25	183.49	181.44 (3; 8.037; 18.907)	**179.57** (3; 5; 9.236; 20.515)	179.74 (3; 5; 8.454; 20.515)	182.91
0.50	144.58	138.31 (3; 8.037; 18.907)	**133.17** (3; 5; 9.236; 20.515)	133.46 (3; 5; 8.454; 20.515)	135.78
0.75	102.35	93.08 (3; 8.037; 18.907)	**86.49** (3; 5; 9.236; 20.515)	86.58 (3; 5; 8.454; 20.515)	90.09
1.00	68.15	58.42 (3; 8.037; 18.907)	52.56 (3; 5; 9.236; 20.515)	**52.34** (3; 5; 8.454; 20.515)	55.62
1.25	44.16	35.82 (3; 8.037; 18.907)	31.59 (3; 5; 9.236; 20.515)	**31.20** (3; 5; 8.454; 20.515)	33.93
1.50	28.51	22.20 (3; 8.037; 18.907)	19.52 (3; 5; 9.236; 20.515)	**19.10** (3; 5; 8.454; 20.515)	21.29
1.75	18.61	14.22 (3; 8.037; 18.907)	12.68 (3; 5; 9.236; 20.515)	**12.26** (3; 5; 8.737; 18.907)	13.69
2.00	12.40	9.54 (3; 8.037; 18.907)	8.65 (3; 5; 9.496; 18.907)	**8.31** (2; 5; 11.021; 20.515)	9.40
2.25	8.49	6.67 (3; 8.577; 17.710)	6.24 (3; 5; 9.496; 18.907)	**5.91** (2; 5; 11.021; 20.515)	6.73
2.50	5.99	4.83 (2; 10.672; 18.907)	4.70 (3; 5; 10.015; 17.710)	**4.41** (2; 5; 11.351; 18.907)	5.07
2.75	4.38	3.66 (2; 10.672; 18.907)	3.60 (2; 3; 11.478; 18.907)	**3.43** (2; 5; 11.351; 18.907)	3.92
3.00	3.31	2.86 (2; 11.342; 17.710)	2.86 (2; 3; 12.112; 17.710)	**2.77** (2; 5; 12.002; 17.710)	3.15

Table 3 ARL profiles for ARL_0=200 and p=10

d	1/1	1/1−m/m	K: r-of-m	CS: r / m	SRR
0.00	200	200	200	200	200
0.25	189.23	187.23 (5; 11.206; 27.722)	186.05 (3; 5; 15.987; 29.588)	**185.99** (3; 5; 14.977; 29.588)	195.09
0.50	161.34	154.85 (5; 11.206; 27.722)	151.18 (3; 5; 15.987; 29.588)	**150.93** (3; 5; 14.977; 29.588)	158.94
0.75	126.15	115.62 (5; 11.206; 27.722)	110.06 (3; 5; 15.987; 29.588)	**109.53** (3; 5; 14.977; 29.588)	113.15
1.00	92.48	80.24 (5; 11.206; 27.722)	74.28 (3; 5; 15.987; 29.588)	**73.52** (3; 5; 14.977; 29.588)	78.65
1.25	64.95	53.33 (4; 12.494; 27.722)	48.16 (3; 5; 15.987; 29.588)	**47.31** (3; 5; 14.977; 29.588)	51.75
1.50	44.53	34.88 (4; 12.494; 27.722)	30.97 (3; 5; 15.987; 29.588)	**30.16** (3; 5; 14.977; 29.588)	33.20
1.75	30.25	22.99 (4; 12.494; 27.722)	20.25 (3; 5; 15.987; 29.588)	**19.56** (3; 5; 14.977; 29.588)	21.71
2.00	20.59	15.40 (3; 14.431; 27.722)	13.69 (3; 5; 15.987; 29.588)	**13.15** (3; 5; 14.977; 29.588)	14.50
2.25	14.17	10.63 (3; 14.431; 27.722)	9.63 (3; 5; 16.320; 27.722)	**9.19** (3; 5; 15.343; 27.722)	10.36
2.50	9.92	7.62 (3; 14.431; 27.722)	7.03 (3; 5; 16.320; 27.722)	**6.68** (2; 5; 18.245; 29.588)	7.67
2.75	7.10	5.61 (3; 15.137; 26.320)	5.36 (3; 5; 16.320; 27.722)	**5.03** (2; 5; 18.245; 29.588)	5.88
3.00	5.21	4.24 (2; 17.808; 27.722)	4.16 (2; 3; 18.815; 27.722)	**3.91** (2; 5; 18.656; 27.722)	4.53

value of the UOCL, then there are numerous combinations of UOCL and UICL yielding a certain ARL_0 value. In such cases, for the selection of a single pair (UICL, UOCL) it is common to pick-up the pair which minimizes the out-of-control ARL at a design (specified) shift d that is considered important enough to be detected quickly. We will refer to this designing method as the optimization method. Optimization methods have been frequently used for the statistical design of control charts supplemented with runs rules (see, e.g., Artiles-Leon et al. 1996; Zhang and Wu 2005; Kim et al. 2009; Lim and Cho 2009 and Acosta-Mejia and Pignatiello 2009). For the design via the optimization method of the CS: r / m control chart, the following steps are suggested:

Step1: Choose the values of p, n, c, r, m and d.
Step2: Minimize the out-of-control ARL at shift d under the constraints $ARL_0 \geq c$ and $CL < UICL < \chi^2_{p;1/c} < UOCL$.

Table 4 provide the optimal values of UICL, UOCL of the CS: r / m control chart (see the homonymous columns) for design shift d=0.50, 1.00, 1.25 and 1.50. We have selected ARL_0=200, 370 and 500, sample size n=1, 2, and 5, and p=5, 10. For the runs rule parameters r, m we examined the cases $2 \leq r < m \leq 5$. At each d value the lowest out-of-control ARL among the examined CS: r / m control charts is given in the “ARL” column along with the runs rule r / m (“CS” column). For comparison purposes we have also provided the corresponding out-of-control ARLs of the standard CSCC (column labeled “1 / 1”).

Table 4 reveals that the reduction achieved in the out-of-control ARL value by the use of the proposed CS: r / m control charts is very attractive, especially for small magnitude shifts and small sample sizes (i.e., for n=1 or 2). As the number of quality characteristics p and/or the in-control ARL value increases, the ARL performance of the proposed charts becomes more superior as compared with the one of the standard CSCC.

In closing this section, we mention that computer programs that produce the numerical results of Table 4 are available from the authors upon request.

4 Conclusions

In the present article we introduce a runs rules based chi-square control chart, the CS: r / m control chart, suitable for monitoring the vector of means of several quality characteristics which are jointly distributed as a multivariate normal distribution. It improves significantly the weak ARL performance of the standard CSCC in the detection of small and/or moderate magnitude shifts in the mean vector, enhancing at the same time its sensitivity in detecting large ones. The ARL performance of the CS: r / m control chart becomes better as the number p of quality characteristics increases. Our numerical experimentation revealed that the CS: r / m control chart can serve as a viable alternative to the standard CSCC, as well as to the CSCC suggested by Aparisi et al. (2004); Khoo et al. (2005) and Koutras et al. (2006).

Finally, in order to assist the practitioners in the selection of the most suitable control scheme we present a practical guidance allowing the easy selection of the optimal combination of r, m, UICL and UOCL which minimizes the out-of-control ARL for a specific shift of the process mean vector. According to this guidance, for the detection of small magnitude shifts in the mean vector we suggest the use of the CS: 3 / 5 control chart while for larger magnitude shifts the CS: 2 / 5 control chart is more suitable.

Acknowledgements The authors would like to thank the referee for his/her useful suggestions and comments which have improved the manuscript.

Table 4 Optimal design parameters and ARLs for the CS: r / m control chart (p=5, 10)

				ARL_0=200					ARL_0=370					ARL_0=500			
p	n	d	CS	UICL	UOCL	ARL	1/1	CS	UICL	UOCL	ARL	1/1	CS	UICL	UOCL	ARL	1/1
5	1	0.50	3/5	8.548	19.800	132.89	144.58	3/5	9.128	21.793	234.00	259.17	3/5	9.409	22.752	308.82	265.74
		1.00	3/5	8.632	19.341	50.93	68.15	3/5	9.194	21.290	79.57	114.28	3/5	9.468	22.229	99.11	344.93
		1.25	3/5	8.702	19.040	30.07	44.16	3/5	9.249	20.958	43.92	71.38	3/5	9.517	21.883	52.93	147.33
		1.50	3/5	8.794	18.714	18.36	28.51	3/5	9.323	20.597	25.20	44.39	3/5	9.584	21.507	29.44	90.42
	2	0.50	3/5	8.575	19.639	92.74	109.20	3/5	9.149	21.617	156.23	190.80	3/5	9.428	22.569	201.84	55.24
		1.00	3/5	8.759	18.827	21.62	33.11	3/5	9.296	20.723	30.29	52.21	3/5	9.559	21.638	35.76	250.85
		1.25	3/5	8.923	18.360	11.53	18.07	3/5	9.428	20.199	14.95	27.02	3/5	9.679	21.090	16.98	65.37
		1.50	2/5	9.145	17.924	6.96	10.28	3/5	9.614	19.694	8.53	14.60	3/5	9.849	20.557	9.41	32.99
	5	0.50	3/5	8.662	19.203	39.63	55.62	3/5	9.218	21.138	59.98	91.68	3/5	9.489	22.071	73.55	17.38
		1.00	3/5	9.230	17.796	6.06	8.66	3/5	9.687	19.543	7.31	12.10	3/5	9.916	20.395	8.01	117.22
		1.25	2/5	11.459	18.607	3.29	4.15	2/5	12.115	20.657	3.82	5.36	2/5	12.444	21.673	4.11	14.30
		1.50	2/4	11.898	17.654	2.13	2.36	2/4	12.565	19.447	2.39	2.84	2/5	12.934	20.323	2.53	6.10
10	1	0.50	3/5	14.982	30.433	150.88	161.34	3/5	15.676	33.144	270.04	292.69	3/5	16.052	34.442	359.25	391.80
		1.00	3/5	14.901	30.319	73.47	92.48	3/5	15.677	33.139	120.58	159.76	3/5	16.050	34.502	135.95	208.91
		1.25	3/5	14.921	30.106	47.29	64.95	3/5	15.686	32.958	73.47	108.86	3/5	16.056	34.342	91.34	140.32
		1.50	3/5	14.956	29.766	30.16	44.53	3/5	15.706	32.621	44.26	72.27	3/5	16.071	34.012	53.52	91.72
	2	0.50	3/5	14.892	30.424	116.42	132.28	3/5	15.674	33.182	202.05	235.76	3/5	16.049	34.508	264.88	312.92
		1.00	3/5	14.942	29.897	35.17	50.78	3/5	15.698	32.754	52.63	83.33	3/5	16.065	34.145	64.25	106.34
		1.25	3/5	15.018	29.269	18.99	29.43	3/5	15.744	32.090	26.26	46.07	3/5	16.101	33.472	30.83	57.49
		1.50	3/5	15.157	28.458	10.97	17.14	3/5	15.835	31.172	14.17	25.58	3/5	16.173	32.514	16.07	31.19
	5	0.50	3/5	14.909	30.234	59.86	78.59	3/5	15.680	33.073	95.78	133.89	3/5	16.052	34.447	120.78	173.92
		1.00	3/5	15.220	28.174	9.35	14.46	3/5	15.878	30.838	11.85	21.24	3/5	16.209	32.160	13.31	25.70
		1.25	2/5	18.435	28.510	4.78	6.70	2/5	19.328	31.078	5.76	9.12	3/5	16.522	30.364	6.27	10.65
		1.50	2/5	18.939	27.058	2.89	3.53	2/5	19.725	29.368	3.32	4.47	2/5	20.118	30.506	3.55	5.04

Appendix

In this section, we present a Markov chain approach suitable for the study of discrete waiting time random variables associated with the time of absorption of a finite Markov chain. Next, we demonstrate that the study of the waiting time distribution of a pattern Λ (simple or compound) can be captured through this approach. As byproduct we obtain the ARL of the CS: 2 / m control chart. Analogous Markov chain techniques for the study of waiting time distributions associated with patterns can be found in Antzoulakos (2001); Fu and Chang (2002); Antzoulakos and Rakitzis (2008), and references there in.

General Results

Let $\{Y_n,\ n=1,\ 2,\ \ldots\}$ be a Markov chain defined on a finite state space $\Omega = \{1,\ 2,\ \ldots,\ k+1\}$ with transition probability matrix $\mathbf{P} = \left[p_{ij}\right]_{(k+1)\times(k+1)}$ and initial probability vector

$$\mathbf{v} = (v_1,\ v_2,\ \ldots, v_{k+1}) = (\Pr(Y_1 = 1),\ \Pr(Y_1 = 2),\ \ldots,\ \Pr(Y_1 = k+1)).$$

Assume that state $k+1$ is the unique absorbing state of the Markov chain. Then, matrix $\mathbf{P}$ can be written in the form

$$\mathbf{P} = \left[\begin{array}{ccc|c} p_{11} & \cdots & p_{1k} & p_{1,k+1} \\ \vdots & \ddots & \vdots & \vdots \\ p_{k1} & \cdots & p_{kk} & p_{k,\,k+1} \\ \hline 0 & \cdots & 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} \mathbf{Q} & (\mathbf{I}-\mathbf{Q})\mathbf{1}' \\ \hline \mathbf{0} & 1 \end{array}\right]_{(k+1)\times(k+1)}.$$

Let T be a waiting time random variable defined by

$$T = \min\{n \geq 1;\quad Y_n = k+1\}.$$

It is then evident that

$$F(n) = \Pr(T \leq n) = \Pr(Y_n = k+1) = \mathbf{v}\mathbf{P}^{n-1}\mathbf{e}'_{k+1},\quad n \geq 1.$$

Since

$$\mathbf{P}^n = \left[\begin{array}{c|c} \mathbf{Q}^n & (\mathbf{I}-\mathbf{Q}^n)\,\mathbf{1}' \\ \hline \mathbf{0} & 1 \end{array}\right]$$

we readily obtain that

$$F(n) = \Pr(T \leq n) = \boldsymbol{\pi}(\mathbf{I} - \mathbf{Q}^{n-1})\mathbf{1}',\quad n \geq 1$$

where $\boldsymbol{\pi} = (\Pr(Y_1 = 1),\ \Pr(Y_1 = 2),\ \ldots,\ \Pr(Y_1 = k))$. The tail probabilities $h(n) = \Pr(T > n)$ and the probability mass function $f(n) = \Pr(T = n)$ of T are given by the following expressions:

$$h(n) = 1 - F(n) = \boldsymbol{\pi}\mathbf{Q}^{n-1}\mathbf{1}',\quad n \geq 1,$$
$$f(n) = h(n-1) - h(n) = \boldsymbol{\pi}\mathbf{Q}^{n-2}(\mathbf{I} - \mathbf{Q})\mathbf{1}',\quad n \geq 2.$$

The generating function $H(s)$ of the tail probabilities and the probability generating function $G(s)$ of T are given, respectively, by

$$H(s) = 1 + \sum_{n=1}^{\infty} h(n)s^n = 1 + \sum_{n=1}^{\infty} \boldsymbol{\pi}\mathbf{Q}^{n-1}1's^n = 1 + s\boldsymbol{\pi}\left(\sum_{n=0}^{\infty}(s\mathbf{Q})^n\right)1 = 1 + s\boldsymbol{\pi}(\mathbf{I} - s\mathbf{Q})^{-1}1',$$
$$G(s) = 1 - (1-s)H_T(s) = s\left(1 - (1-s)\boldsymbol{\pi}(I - s\mathbf{Q})^{-1}1'\right).$$

Furthermore, by exploiting the well known formula

$$\frac{d}{ds}(\mathbf{I} - s\mathbf{Q})^{-m} = m(\mathbf{I} - s\mathbf{Q})^{-(m+1)}\mathbf{Q},\ m \geq 1,$$

we deduce that the descending factorial moments $\mu'_{[m]}$, $m \geq 1$, of T are given by

$$\mu'_{[m]} = m\frac{d^{m-1}}{ds^{m-1}}H_T(s)\bigg|_{s=1} = \begin{cases} 1 + \pi(\mathbf{I} - \mathbf{Q})^{-\mathbf{1}'}, & m = 1 \\ m!\pi(I - \mathbf{Q})^m\mathbf{Q}^{m-2}\mathbf{1}', & m \geq 2. \end{cases}$$

Waiting Time for a Compound Pattern

Let $\{X_n, n \geq 1\}$ be a sequence of independent and identically distributed rv's taking values in a set $A = \{a_1, a_2, ..., a_s\}$, $s \geq 2$, with probabilities $p_j = \Pr(X_n = a_j)$ ($n \geq 1$, $1 \leq j \leq s$, $\Sigma p_j = 1$). A simple pattern is defined to be a finite sequence of outcomes from A and a compound pattern, say $\Lambda = \cup_{i=1}^{m}\Lambda_i$, is defined to be the union of m non-overlapping distinct simple patterns. Let T be the minimum number of trials required to obtain the pattern Λ.

Denote by ℓ_i, $1 \leq i \leq m$, the length of the simple pattern Λ_i and let $S(\Lambda_i)$ be the set containing all the initial subpatterns of Λ_i with lengths greater than 1 and less than ℓ_i. Set $C = \sum_{i=1}^{m} S(\Lambda_i)$ and let B be the set of all the patterns of length 1 that can be formed from the set A that do not belong to the compound pattern Λ. For example, for $A = \{1, 2, 3\}$, $\Lambda_1 = 11232$, $\Lambda_2 = 3$ and $\Lambda = \Lambda_1 \cup \Lambda_2 = \{11232, 3\}$, we have that $S(\Lambda_1) = \{11, 112, 1123\}$, $S(\Lambda_2) = \emptyset$ and $B = \{1,2\}$.

Set $D = B \cup C \cup \Lambda$, assume that $|B \cup C| = k$, assign to each pattern in $B \cup C$ a unique label (number) from 1 to k and correspond to each simple pattern of Λ the label $k + 1$. We define a Markov chain $\{Y_n, n \geq 1\}$ with state space

$$\Omega = \{1,\ 2,\ ...,\ k,\ k+1\}$$

operating parallel to $\{X_n, n \geq 1\}$ according to the following two conditions: (a) the state $k + 1$ is an absorbing state in which the Markov chain enters for the first time when the pattern Λ occurs in $X_1, X_2, ...$, and (b) we assign to Y_n the value j ($1 \leq j \leq k$) if the longest subpattern in D that matches with the ending pattern in $X_1, X_2, ..., X_n$, counting backward, is identified to be the pattern corresponding to the label j.

The above definitions establish a time homogeneous Markov chain on Ω with initial probability vector

$$\mathbf{v} = (v_1,\ v_2,\ \ldots, v_{k+1}) = (\Pr(Y_1 = 1),\ \Pr(Y_1 = 2),\ ...,\ \Pr(Y_1 = k+1))$$

and transition probability matrix $\mathbf{P} = \left[p_{ij}\right]_{(k+1)\times(k+1)}$ defined by

$$p_{ij} = \sum p_r,\quad 1 \leq i \leq k,\ \ 1 \leq j \leq k+1,$$

where the sum is taken over all p_r corresponding to a_r for which the ending pattern "i" changes to the ending pattern "j". It is obvious that

$$\Pr(T \leq n) = \Pr(Y_n = k+1), \quad n \geq 1$$

and therefore the study of the waiting time random variable T may be performed by exploiting the results established in A1.

Application to the CS: *r* / *m* Control Chart

Here, we will demonstrate how the results of A1 and A2 may be applied for study of the CS: r / m control chart. Even though we focus only on the derivation of the ARL, our results are appropriate for a complete study of the run length distribution, such as the computation of the percentile values which may be serve as an alternative measure of the performance of a manufacturing process (see, e.g., Montgomery (2005) and Palm (1990)). In the sequel we confine ourselves to the study of the CS: 2 / m control since the analysis for arbitrary values of r, m may be performed in a similar manner.

Let $\{X_n,\ n \geq 1\}$ be a sequence of independent and identically distributed rv's taking values in the set A={0,1,2,3} with probabilities $p_j = \Pr(X_n = j)$ (n≥1, 0≤j≤3) given by (2.1). We introduce the simple patterns

$$\Lambda_1 = 3, \quad \Lambda_2 = 22, \quad \Lambda_3 = 212, \quad \Lambda_4 = 2112, \ \ldots, \ \Lambda_m = 2\underbrace{11\cdots1}_{m-2}2, \quad m \geq 2.$$

and denote by T the waiting time for the occurrence of the compound pattern $\Lambda = \cup_{i=1}^{m}\Lambda_i$.

It follows from the above set-up that the run length distribution of the CS: 2 / m control chart coincides with the waiting time distribution T of the compound pattern Λ. For the study of T we employ the results established in A1 and A2. We define a Markov chain $\{Y_n,\ n \geq 1\}$ with state space Ω={1, 2, ...,m + 2}, where the states 1, 2, ..., m+1 correspond to the patterns "0", "1", "2", "21", "211", ... , "$2\underbrace{11\cdots1}_{m-2}$", respectively, and state m + 2 is an absorbing state corresponding to pattern Λ. The transition probability matrix $\mathbf{P}$ of the Markov chain is given by

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{Q} & (\mathbf{I}-\mathbf{Q})\mathbf{1}' \\ \hline \mathbf{0} & 1 \end{array}\right] = \left[\begin{array}{cccccccc|c} p_0 & p_1 & p_2 & 0 & 0 & \cdots & 0 & 0 & p_3 \\ p_0 & p_1 & p_2 & 0 & 0 & \cdots & 0 & 0 & p_3 \\ p_0 & 0 & 0 & p_1 & 0 & \cdots & 0 & 0 & p_2+p_3 \\ p_0 & 0 & 0 & 0 & p_1 & \cdots & 0 & 0 & p_2+p_3 \\ p_0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & p_2+p_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ p_0 & 0 & 0 & 0 & 0 & \cdots & 0 & p_1 & p_2+p_3 \\ p_0 & p_1 & 0 & 0 & 0 & \cdots & 0 & 0 & p_3 \\ \hline 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{array}\right]_{(m+2)\times(m+2)}$$

and $\boldsymbol{\pi} = (\Pr(Y_1 = 1),\ \Pr(Y_1 = 2),\ \ldots,\ \Pr(Y_1 = m+1)) = (p_0,\ p_1,\ p_2,\ 0,\ \ldots,\ 0)_{1\times(m+1)}$.

Carrying out some algebra, we may establish the next formula for the ARL of the CS:2 / m control chart,

$$\begin{aligned}\mathrm{ARL} &= E(T) = 1 + \boldsymbol{\pi}(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{1}' \\ &= \frac{1-p_1+p_2\left(1-p_1^{m-1}\right)}{(1-p_1)\left(1-p_0-p_1\left(1+p_2p_1^{m-2}\right)\right)-p_0p_2\left(1-p_1^{m-1}\right)}.\end{aligned}$$

In closing we mention that the proposed methodology can be likewise utilized, after some trivial modifications, for the study of the runs rules based control charts proposed by Khoo et al. (2005) and Koutras et al. (2006). The details are left to the reader.

References

Acosta-Mejia C, Pignatiello JJ (2009) ARL-design of S charts with k-of-k runs rules. Commun Stat-Simul C 38:1625–1639

Alt FB, Smith ND (1988) Multivariate process control. In: Krishnaiah PR, Rao CR (eds) Handbook of Statistics vol. 7. Elsevier, Amsterdam, pp 333–351

Antzoulakos DL (2001) Waiting times for patterns in a sequence of multistate trials. J Appl Probab 38:508–518

Antzoulakos DL, Rakitzis AC (2008) The modified r out of m control chart. Commun Stat-Simul C 37:396–408

Aparisi F, Champ CW, Diaz JCG (2004) A performance analysis of Hotelling's χ^2 control chart with supplementary runs rules. Qual Eng 16(3):13–22

Artiles-Leon N, David HT, Meeks HD (1996) Statistical optimal design of control charts with supplementary stopping rules. IIE Trans 28(3):225–236

Bersimis S, Psarakis S, Panaretos J (2007) Multivariate statistical process control charts: an overview. Qual Reliab Eng Int 23:517–543

Crosier RB (1988) Multivariate generalizations of cumulative sum quality-control schemes. Technometrics 30:291–303

Fu JC, Chang YM (2002) On probability generating functions for waiting time distributions of compound patterns in a sequence of multistate trials. J Appl Probab 39:70–80

Hotelling H (1947) Multivariate quality control. In: Eisenhart C, Hastay M, Wallis WA (eds) Techniques of statistical Analysis. McGraw-Hill, New York, pp 111–184

Khoo MBC (2005) A nonrigorous approach of incorporating sensitizing rules into multivariate control charts. J Mod Appl Stat Methods 4(2):567–582

Khoo MBC, Quah SH (2003) Incorporating runs rules into Hotelling's χ^2 control charts. Qual Eng 15 (4):671–675

Khoo MBC, Quah SH (2004) Alternatives to the multivariate control chart for process dispersion. Qual Eng 16(3):423–435

Khoo MBC, Quah SH, Low HC (2005) Powerful rules for the Hotelling's χ^2 control chart. Qual Eng 17:139–149

Kim Y-B, Hong J-S, Lie C-H (2009) Economic-statistical design of 2-of-2 and 2-of-3 runs rule scheme. Qual Reliab Eng Int 25:215–228

Koutras MV, Bersimis S, Antzoulakos DL (2006) Improving the performance of the chi-square control chart via runs rules. Methodol Comput Appl 8:409–426

Koutras MV, Bersimis S, Maravelakis PE (2007) Statistical process control using Shewhart control charts with supplementary runs rules. Methodol Comput Appl 9:207–224

Lim T-J, Cho M (2009) Design of control charts with m-of-m runs rules. Qual Reliab Eng Int doi:10.1002/qre.1023

Lowry CA, Woodall WH, Champ CW, Rigdon SE (1992) A multivariate exponentially weighted moving average control chart. Technometrics 34:46–53

Montgomery DC (2005) Introduction to statistical quality control. Wiley, New York

Palm AC (1990) Tables of run length percentiles for determining the sensitivity of Shewhart control charts for averages with supplementary runs rules. J Qual Technol 22:289–298

Pignatiello JJ Jr, Runger GC (1990) Comparisons of multivariate CUSUM charts. J Qual Technol 22:173–186

Zhang S, Wu Z (2005) Designs of control charts with supplementary runs rules. Comput Ind Eng 49:76–97