RRPA91070499 (13 .P)

附件一

行政院國家科學委員會補助專題研究計畫 ☑成果報告 □期中進度報告

適用於計數值管制圖的一組新的管制界限

計畫類別：☑ 個別型計畫 □ 整合型計畫

計畫編號：NSC －91－2118－M－032－007

執行期間： 91 年 8 月 1 日至 92 年 7 月 31 日

計畫主持人：蔡宗儒

共同主持人：

計畫參與人員： 江宜蓁(碩士生兼任助理)

成果報告類型(依經費核定清單規定繳交)：☑精簡報告 □完整報告

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：淡江大學

中 華 民 國 92 年 9 月 5 日

Improved Square Root Transformation to Adjust Control Limits on Attribute Charts

TZONG-RU TSAI

Department of Statistics, Tamkang University, Tamsui, Taipei, Taiwan

摘要

長久以來我們都是使用常態近似來發展標準的計數值管制圖，但對於某些樣本數及參數值所造成的母體機率分配呈現偏態分布時，常態近似就不適合被使用來發展計數值管制圖。本文提供一個改良的平方根轉換 −ISRT，來建造ISRT $p-$圖及ISRT $np-$圖以監測二項資料。而且，當我們要監測波氏資料時，我們可以將波氏近似使用在ISRT $np-$圖上來得到 ISRT $c-$圖。比較ISRT $p-$圖，迴歸基準的$p-$圖，Arcsine $p-$圖，修正的$p-$圖，$Q-$圖及古典的$3-\sigma$ $p-$圖，可以發現ISRT $p-$圖，迴歸基準的$p-$圖及Arcsine $p-$圖在所需的最小樣本數以保證下管制界限是有效的要求下可以比其他方法需要較少的樣本數。此外，模擬的結果顯示，當參數未知時，ISRT管制圖可配合任何特定的真實管制界限的連串長度分配的百分比。

Standard attribute control charts have been developed historically using the normal approximation, but for some sample size and process parameter values, the normal approximation is far from adequate, mainly due to skewness in the exact distribution. In this paper, an improved square root transformation, named ISRT, is used to construct the ISRT p-chart and np-chart for charting the binomial data. Moreover, the ISRT c-chart can be obtained from an ISRT np-chart using the Poisson approximation for charting the Poisson data. Compared with the ISRT p-chart, the regression-based p-chart, the Arcsine p-chart, the modified p-chart, the Q-chart and the classical 3-sigma p-chart, the minimum sample sizes necessary such that the lower control limits are effective for the ISRT p-chart, the regression-based p-chart and the Arcsine p-chart are smaller than others. Moreover, simulation results indicate that the ISRT control

charts can match any specific percentile point of run length distribution of the true limits when the parameter is unknown.

Keywords: Attribute control charts, classical normal approximation, control limits, improved square root transformation, Q-chart.

1 Introduction

Attribute control charts have been historically developed using the normal approximation. Suppose the process observations are taken from a binomial distribution with parameters n and p, where n denotes the items produced and inspected and p denotes the proportion of nonconforming items. The classical 3-sigma p-chart and np-chart have been developed for charting the binomial data. When p is known, the classical 3-sigma p-chart and np-chart can be constructed with the center lines (CLs) p and np and the control limits $p \pm 3\sqrt{p(1-p)/n}$ and $np \pm 3\sqrt{np(1-p)}$, respectively.

The normal approximation to binomial distribution can perform adequately if both conditions $np \geq 5$ and $n(1-p) \geq 5$ are fulfilled. Schader and Schmid (1989) showed that the normal approximation to binomial distribution performs poorly even when the often-used rules of thumb such as $np \geq 5$ are met and the accuracy of the approximation is shown to depend heavily upon the value of p. Ryan and Schwertman (1997) showed that the upper tail probability in the classical 3-sigma np-chart is usually too large and the lower tail probability is usually too small, especially when p is small. This inadequacy in approximation is mainly due to the skewness in the exact distribution. For improving performances of the classical 3-sigma p-chart and np-chart, some alternative approximation rules were provided so that tail areas of the charts can be close to those of the exact distribution.

In Section 2, we will describe four different charting approaches in the literature. In Section 3, we will introduce a new approach called the improved square root transformation (ISRT). In Section 4, some numerical results are presented to assess the performances of these approaches. Some conclusions are made in Section 5.

2 Recent Approaches

In this section, we will introduce four different methods which can improve the performances of tail probabilities in a p-chart or an np-chart. The first one is the Arcsine p-chart. Suppose that p is known. Define

$$Y_i = 2\sqrt{n_i}\left[\sin^{-1}\sqrt{\frac{x_i + 3/8}{n_i + 3/4}} - \sin^{-1}\sqrt{p}\right],$$

where x is the number of nonconforming items in a sample. It can be shown that Y_i is approximately standard normal. Hence, we can plot $Y_1, Y_2, \ldots,$ on a chart with the center line $CL = 0$ and upper control limit at $UCL = 3$ and lower control limit at $LCL = -3$. When p is unknown, we can use $\hat{p}_i = \dfrac{x_1 + x_2 + \cdots + x_i}{n_1 + n_2 + \cdots + n_i}$ to estimate p, and then define

$$Y_i = 2\sqrt{n_i}\left[\sin^{-1}\sqrt{\frac{x_i + 3/8}{n_i + 3/4}} - \sin^{-1}\sqrt{\hat{p}_{i-1}}\right],$$

for $i \geq 2$. The Arcsine p-chart has two beneficial effects: (1) the tail areas will be close to the tail areas under normality, and (2) for a given p, the minimum sample size necessary to produce a positive LCL is much smaller than the sample size required using the classical 3-sigma p-chart (see Ryan and Schwertman (1997)).

The second approach is the Q-chart proposed by Quesenberry (1991a, 1991b, 1991c). Suppose that p is known. Let $u_i = B(x_i; n_i, p)$ be the probability $Pr(X_i \leq x_i)$ for a binomial random variable X_i. Quesenberry (1991a) suggested that the statistics $Q_i = \Phi^{-1}(u_i)$, $i = 1, 2, \ldots,$ are plotted on a chart with $CL = 0$, $UCL = 3$, and $LCL = -3$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. In practice, p is usually unknown. We can define $s_i = \sum_{j=1}^{i} n_j$, $t_i = \sum_{j=1}^{i} x_j$ and let $u_i = H(x_i; t_i, n_i, s_{i-1})$ be the probability $Pr(X_i \leq x_i)$ for a hypergeometric random variable X_i, and $Q_i = \Phi^{-1}(u_i)$. Then, we can plot $Q_1, Q_2, \ldots,$ on a chart with center line $CL = 0$, and control limits $UCL = 3$, and $LCL = -3$.

Quesenberry (1991b) indicated that, in general, the Q-chart has tail probability less than 0.00135 for the lower tail and more than 0.00135 for the upper tail. Moreover, Quesenberry (1991b) concluded that the Arcsine p-chart gives a better approximation to the nominal lower tail area, and the Q-chart provides a better approximation to the nominal upper tail area. However, both Arcsine p-chart and Q-chart plot a transformed statistic rather than the statistic of interest; and computations for both transformed statistics are complicated.

Winterbottom (1993) and Chen (1998) used the Cornish-Fisher expansion of quantiles to construct a modified p-chart. For a given p, we can plot $w_i = x_i/n_i$, for $i = 1, 2, \ldots$, on a chart with $CL = p$ and control limits

$$UCL = p + 3\sqrt{\frac{p(1-p)}{n_i}} + \frac{4}{3}\left(\frac{1-2p}{n_i}\right),$$
$$LCL = p - 3\sqrt{\frac{p(1-p)}{n_i}} + \frac{4}{3}\left(\frac{1-2p}{n_i}\right).$$

If p is unknown, we can plot w_i's on a chart with center line $CL = \hat{p}_{i-1}$ and control limits

$$UCL = \hat{p}_{i-1} + 3\sqrt{\frac{\hat{p}_{i-1}(1-\hat{p}_{i-1})}{n_i}} + \frac{4(1-2\hat{p}_{i-1})}{3n_i},$$
$$LCL = \hat{p}_{i-1} - 3\sqrt{\frac{\hat{p}_{i-1}(1-\hat{p}_{i-1})}{n_i}} + \frac{4(1-2\hat{p}_{i-1})}{3n_i},$$

where $\hat{p}_i = \dfrac{x_1 + x_2 + \ldots + x_i}{n_1 + n_2 + \ldots + n_i}$. Moreover, Winterbottom (1993) used the Cornish and Fisher expansion of quantiles to develop a modified c-chart for charting the Poisson data. The center line of modified c-chart is $CL = c$ and its control limits are

$$UCL = c + 3\sqrt{c} + \frac{4}{3},$$
$$LCL = c - 3\sqrt{c} + \frac{4}{3},$$

where c is the mean of the Poisson distribution and it can be either specified or replaced by an estimate $\bar{c}$ computed using an in-control baseline data.

Ryan and Schwertman (1997) provided a regression-based np-chart with center line

$$CL = np \tag{1}$$

and control limits

$$UCL = 0.6195 + 1.0052np + 2.983\sqrt{np}, \tag{2}$$

$$LCL = 2.9529 + 1.01956np - 3.2729\sqrt{np}. \tag{3}$$

Ryan and Schwertman (1997) suggested that the control limits computed from equations (2) and (3) should be rounded to the nearest integer. By dividing each of the equations (1), (2), and (3) by n, control limits of the regression-based p-chart can be obtained. The advantages of the regression-based p-chart and np-chart are the computation on equations (1), (2), and (3) are simple and they avoid the cumbersome expressions found in some confidence limits approximations. Replacing np by c to equations (1), (2), and (3), a regression-based c-chart can be constructed. The value of p can be either specified or replaced by the estimate $\bar{p}$ computed from an in-control baseline data.

3 ISRT Attribute Control Charts

Let X be a binomial random variable with parameters n and p, and let $\hat{p} = \dfrac{X}{n}$ be the sample proportion. If p is small, the normal approximation to binomial distribution is inadequate, mainly due to skewness in the exact distribution. To overcome this defect, a square root transformation, named the ISRT, is used to construct an ISRT p-chart.

Suppose that $g(\hat{p}) = \sqrt{\hat{p}}$. Using the Taylor series expansion for g to the second order, we have

$$g(\hat{p}) \cong g(p) + g'(p)(\hat{p} - p) + \frac{g''(p)}{2}(\hat{p} - p)^2.$$

Equivalently,

$$\sqrt{n}\left[g(\hat{p}) - g(p) - \frac{g''(p)}{2}(\hat{p} - p)^2\right] \cong g'(p)\sqrt{n}(\hat{p} - p).$$

Hence, both $\sqrt{n}[g(\hat{p}) - g(p) - \frac{g''(p)}{2}(\hat{p} - p)^2]$ and $g'(p)\sqrt{n}(\hat{p} - p)$ have the same limiting distribution; that is, $\sqrt{n}[g(\hat{p}) - g(p) - \frac{g''(p)}{2}(\hat{p} - p)^2]$ is asymptotically normally distributed with mean 0 and variance $[g'(p)]^2 p(1-p)$. Then, the 3-sigma ISRT p-chart with $CL = g(p)$ and control limits are

$$UCL_p = g(p) + 3|g'(p)|\sqrt{\frac{p(1-p)}{n}} + \frac{g''(p)}{2}\, e^2, \tag{4}$$

$$LCL_p = g(p) - 3|g'(p)|\sqrt{\frac{p(1-p)}{n}} + \frac{g''(p)}{2}\, e^2, \tag{5}$$

where $e = |\hat{p} - p|$ is the absolutely sampling error in process. The advantage of using equations (4) and (5) to construct p-chart is that the extra term $\frac{g''(p)}{2}e^2$ can adjust the control limits so that the approximating distribution is closer to the exact distribution. When p is small, the binomial distribution is positively skewed and has a long tail to the right. Hence, we can take the sampling error e to be $3\sqrt{\frac{p(1-p)}{n}}$ in the LCL_p, and take e to be $2\sqrt{\frac{p(1-p)}{n}}$ in the UCL_p. Then, the center line and control limits of an ISRT p-chart can be expressed explicitly as

$$CL_p = \sqrt{p}, \tag{6}$$

$$UCL_p = \sqrt{p} + \frac{3}{2}\sqrt{\frac{1-p}{n}} - \frac{1}{2}\left(\frac{1-p}{n\sqrt{p}}\right), \tag{7}$$

$$LCL_p = \sqrt{p} - \frac{3}{2}\sqrt{\frac{1-p}{n}} - \frac{9}{8}\left(\frac{1-p}{n\sqrt{p}}\right). \tag{8}$$

Moreover, the ISRT np-chart can be developed by multiplying $\sqrt{n}$ to equations (6), (7) and (8) and, hence the center line and the control limits are

$$\begin{aligned} CL_{np} &= \sqrt{np}, \\ UCL_{np} &= \sqrt{np} + \frac{3}{2}\sqrt{1-p} - \frac{1}{2}\left(\frac{1-p}{\sqrt{np}}\right), \\ LCL_{np} &= \sqrt{np} - \frac{3}{2}\sqrt{1-p} - \frac{9}{8}\left(\frac{1-p}{\sqrt{np}}\right). \end{aligned}$$

Using the Poisson approximation to binomial, the ISRT c-chart can also be developed. If the sample size n is large and p is small and $np \to c$ and $(1-p) \to 0$, the center line and control limits of a ISRT c-chart can be written as

$$\begin{aligned} CL_c &= \sqrt{c}, \\ UCL_c &= \sqrt{c} + \frac{3}{2} - \frac{1}{2}\left(\frac{1}{\sqrt{c}}\right), \\ LCL_c &= \sqrt{c} - \frac{3}{2} - \frac{9}{8}\left(\frac{1}{\sqrt{c}}\right). \end{aligned}$$

4 Numerical Study

In practical control chart applications for the binomial data, the values of p are often small such that the binomial distribution is highly skewed. Hence, any attempt for charting the value of p with symmetric control limits is subject to making more false alarms in detecting an increase or a decrease in p than claimed. In this paper, three criteria are provided for discussing the performance of the ISRT p-chart. They are: (1) the minimum value of the sample size, denoted by n^*, such that the LCL is effective, (2) the closeness of the false alarm probabilities to the nominal values for both over UCL and under LCL, and (3) how close the chart can match to a specified percentile point of run length (RL) distribution when the parameter is unknown.

First, we compare the regression-based p-chart, the Arcsine p-chart, the modified p-chart, the Q-chart, the classical 3-sigma p-chart with the ISRT p-chart based on the value of n^*. For the Arcsine p-chart and the Q-chart, the effective lower control limit means that $P(Y_i < LCL) > 0$ and $P(Q_i < LCL) > 0$, respectively; moreover, it means that $LCL > 0$ for others. Table 1 shows the values of n^* for these charts when the value of p is smaller than 0.1. We can find that values of n^* for the ISRT p-chart, the regression-based p-chart and the Arcsine p-chart are smaller than others. Likewise, for the classical c-chart, the value of LCL is positive if $c > 9$; for the regression-based c-chart, the value of LCL is positive if $c > 4.07$; for the modified c-chart, the value of LCL is positive if $c > 6.04$ and for the ISRT c-chart, the value of LCL is positive if $c > 4.20$.

Next, the modified c-chart and the regression-based c-chart are selected to compare the closeness of the false alarm probabilities to the nominal values for both over UCL and under LCL with the ISRT c-chart. Let Z_i denote any one of the three plotting statistics on these charts and the false alarm probabilities $P(Z_i > UCL)$ and $P(Z_i < LCL)$ are computed for $c = 4$ to 25. Because the Poisson distribution is discrete, it is hard to require that both exact tail areas equal to 0.00135. For a given value of c, we define the exact c-chart as the one whose α_L (the lower tail area under LCL) and α_U (the upper tail area over the UCL) are as close to 0.00135 as possible, but are not 50% larger than 0.00135; that is, both α_L and α_U are restricted to be less than or equal to $0.00315(1 + 0.5) = 0.002025$. The numerical results are given in Table 2. We see that the differences among three c-charts are very small and all of them agree with the exact c-chart. Likewise, the simulation results provided by Lin (2002) for the p-chart with parameters $p = 0.1$ and $n = 10(10)200$, and $p = 0.01$ and $n = 100(100)2000$, also showed that all the modified p-chart, regression-based p-chart and the ISRT p-chart agree with the exact p-chart.

The above discussion are made under the assumption that p or c are known, but the values of p and c are often unknown in practical applications. For evaluating the performance of the charts when the parameters are unknown, how close the control limits of the three charts can match to a specific percentile point of RL distribution for the true limits is evaluated. Assume that the process is in-control and m initial subgroups of the binomial data with parameters n and p are observed, and then k future subgroup observations are obtained immediately after m initial subgroup observations. Let γ be the desired probability of a signal within observations $m+1, m+2, \ldots, m+k$. Then

$$P(LCL < Z_i < UCL, i = m+1, m+2, \ldots, m+k) = 1 - \gamma.$$

For the estimated limits which perform similarly to the true limits, γ is set to be equal to the corresponding RL distribution percentile with the true limits. That is,

$$\gamma = P(RL \leq k) = 1 - (1-\alpha)^k,$$

where α is the false alarm probability for one subgroup. A benchmark for the probability is $1 - (1 - 0.0027)^k$; for example, if $k = 10$, the benchmark probability is 0.0267. However, when p is unknown, $Z_{m+1}, Z_{m+2}, \ldots, Z_{m+k}$ are dependent and the probability γ depends upon the initial subgroup size m. The correlation among Z_i's vanishes if m is large enough. Nevertheless, we can adopt the value of γ as measures of the false alarm probability of the three charts and estimate them.

In our simulation study, the future k subgroup observations are generated one at a time, and the sample proportion on each future subgroup were plotted on each chart until a false alarm was issued or until all k subgroups were plotted. This procedure was replicated 50000 times and the probability of a signal within k subgroups was estimated. The estimated probabilities for the ISRT np-chart, the regression-based np-chart and the modified np-chart are listed in Table 3. It can be seen that the control limits of the ISRT np-chart perform similarly to the true limits for almost all combinations of n, p and k, and the control limits of the regression-based np-chart produces a larger false alarms than the true limits do when the value of p is unknown. The false alarm probability of the modified np-chart becomes large when the value of k is large or the value of p is small. The numerical results recommend that the ISRT np-chart is adequate for charting the binomial data when the parameter p is unknown.

5 Conclusions

This paper provides three new attribute control charts based on the improved square root transformation. They are the ISRT p-chart, ISRT np-chart and ISRT c-chart. The false alarm probabilities of these charts are close to the nominal values whenever the parameter is known or unknown. The numerical results indicate that the ISRT p-chart and the ISRT c-chart almost coincides with the exact p-chart and exact c-chart, respectively, under reasonable restrictions. Moreover, the ISRT np-chart can match any specific percentile point of run length distribution of the true limits when the parameter is unknown. Though the ISRT attribute control chart plots a square root transformed statistic rather than the statistic of interest, the link between the plotted statistic and the statistic of interest is simple to be interpreted. In practical application, the ISRT attribute control charts are recommended for charting the binomial and the Poisson data due to its simplicity and good performance.

參考書目

Chen, G. (1998) An improved p chart through simple adjustments, *Journal of Quality Technology*, 30, pp. 142-151.

Lin, C.-C. (2002) Improved Square Root Transformation for Attribute Control Charts, Master thesis, Tamkang University, Department of Statistics (in Chinese).

Quesenberry, C. P. (1991a) SPC Q charts for start-up process and short or long runs, *Journal of Quality Technology*, 23, pp. 213-224.

Quesenberry, C. P. (1991b) SPC Q charts for a binomial parameter: Short or long runs, *Journal of Quality Technology*, 23, pp. 239-246.

Quesenberry, C. P. (1991c) SPC Q charts for a Poisson parameter λ: Short or long runs, *Journal of Quality Technology*, 23, pp. 296-303.

Ryan, T. P. & Schwertman, N. C. (1997) Optimal limits for attributes control charts, *Journal of Quality Technology*, 29, pp. 86-98.

Schader, M. & Schmid, F. (1989) Two rules of thumb for the approximation of the binomial distribution by the normal distribution, *The American Statistician*, 43, pp. 23-24.

Winterbottom, A. (1993) Simple adjustments to improve control limits on attribute charts, *Quality and Reliability Engineering International*, 9, pp. 105-109.

Table 1: Minimum value of sample size such that the LCL is effective

chart	p					
	0.1	0.05	0.02	0.01	0.005	0.001
ISRT	38	80	206	416	836	4195
regression-based	41	82	204	408	815	4072
Arcsine	43	88	222	445	891	4461
modified	58	119	300	602	1206	6037
Q-chart	63	129	328	658	1319	6605
classical	81	171	441	891	1791	8991

Table 2: False alarm probabilities for four different c-charts

c	under LCL				over UCL			
	modified	regression based	ISRT	exact	modified	regression based	ISRT	exact
4	NA	NA	NA	NA	0.00092	0.00284	0.00284	0.00092
5	NA	0.00674	0.00674	NA	0.00070	0.00545	0.00202	0.00202
6	NA	0.00248	0.00248	NA	0.00140	0.00363	0.00140	0.00140
7	0.00091	0.00091	0.00091	0.00091	0.00096	0.00241	0.00241	0.00096
8	0.00034	0.00302	0.00034	0.00034	0.00159	0.00372	0.00159	0.00159
9	0.00123	0.00123	0.00123	0.00123	0.00106	0.00243	0.00243	0.00106
10	0.00050	0.00277	0.00050	0.00050	0.00159	0.00345	0.00159	0.00159
11	0.00121	0.00121	0.00121	0.00121	0.00104	0.00225	0.00225	0.00104
12	0.00052	0.00229	0.00052	0.00052	0.00147	0.00305	0.00147	0.00147
13	0.00105	0.00105	0.00105	0.00105	0.00097	0.00397	0.00199	0.00199
14	0.00181	0.00181	0.00047	0.00181	0.00131	0.00261	0.00131	0.00131
15	0.00086	0.00279	0.00086	0.00086	0.00172	0.00331	0.00172	0.00172
16	0.00138	0.00138	0.00040	0.00138	0.00113	0.00219	0.00219	0.00113
17	0.00067	0.00206	0.00067	0.00067	0.00145	0.00273	0.00145	0.00145
18	0.00104	0.00104	0.00104	0.00104	0.00096	0.00333	0.00181	0.00181
19	0.00151	0.00151	0.00052	0.00151	0.00121	0.00223	0.00223	0.00121
20	0.00078	0.00209	0.00078	0.00078	0.00149	0.00269	0.00149	0.00149
21	0.00111	0.00111	0.00111	0.00111	0.00100	0.00320	0.00181	0.00181
22	0.00150	0.00150	0.00058	0.00150	0.00121	0.00216	0.00121	0.00121
23	0.00081	0.00198	0.00081	0.00198	0.00146	0.00255	0.00146	0.00146
24	0.00108	0.00108	0.00108	0.00108	0.00099	0.00298	0.00173	0.00173
25	0.00142	0.00142	0.00059	0.00142	0.00118	0.00204	0.00204	0.00118

Note: NA denotes 'not available'.

Table 3: Performance assessment for $\alpha = 0.0027$

n	p	k	modified	regression based	ISRT	γ
150	0.10	10	0.02876	0.06008	0.02966	0.02667
300	0.10	10	0.02922	0.03864	0.02664	0.02667
150	0.10	20	0.05492	0.11662	0.05620	0.05263
300	0.10	20	0.05704	0.07488	0.05286	0.05263
150	0.10	30	0.08102	0.16910	0.08276	0.07790
300	0.10	30	0.08328	0.11190	0.07810	0.07790
450	0.05	10	0.03988	0.07312	0.03250	0.02667
600	0.05	10	0.03938	0.06320	0.03292	0.02667
450	0.05	20	0.07518	0.14040	0.06390	0.05263
600	0.05	20	0.07206	0.11586	0.06132	0.05263
450	0.05	30	0.11482	0.20650	0.09730	0.07790
600	0.05	30	0.10512	0.16882	0.08932	0.07790
2000	0.01	10	0.04308	0.09312	0.03502	0.02667
2000	0.01	20	0.08174	0.17378	0.06734	0.05263
2000	0.01	30	0.12082	0.25072	0.10222	0.07790