Thailand Statistician
January 2012; 10(1) : 107-120
http://statassoc.or.th
Contributed paper

Control Charts for Zero-Inflated Binomial Models

Bunpen Yawsaeng and Tidadeaw Mayureesawan*
Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand.
* Author for correspondence; e-mail: tidadeaw@yahoo.com

Received: 16 November 2011
Accepted: 7 March 2012

Abstract

In this paper, we study the performance of control charts when zero-inflated counts are observed in a binomial distribution. The control charts are based on different binomial confidence interval methods, namely, np_{ZIB}-chart by ZIB model, np_J-chart by Jeffreys prior interval, np_W-chart by Wilson interval, np_{AC}-chart by Agresti-Coull interval, np_{BS}-chart by Blyth-Still interval and compared with the traditional Shewhart control chart, np-chart. The proportions of observed zero (φ) are studied in 0.3(0.1)0.9 and variances (σ^2) are studied in 2(1)5 with a different magnitude of the shift in the nonconforming proportion (θ).The Average Run Length (ARL) and the Average Coverage Probability (ACP) are used as performance indicators. The control charts signal an out-of-control process when K consecutive points exceed an upper control limit, K = 1, 2, 3, 4, 5. The result shows that when the process is in-control, the np_W-chart is more preferable if K is low (K = 1 and 2) and φ is high (0.9) for all levels of σ^2. The np-chart performs better than the other control charts if K is high (K = 3, 4 and 5) and φ is also high (0.8-0.9) in all levels of σ^2. For the out-of-control process, when K = 2, 3 and 4, the np_j-chart has a higher efficiency than the other control charts if σ^2 and φ are relatively low (σ^2 = 2 and φ = 0.3) in all levels of θ. Moreover, the np-chart is more fitting when K = 2, 3, 4 and 5 and φ is considerably high (0.9) in all levels of σ^2 and θ.

Keywords: average of coverage probability, average run length, control rules, zero-inflated binomial model.

1. Introduction

Attribute control charts arise in the quality control of a manufacturing process where the quality of the individual items of a product is checked against the product specifications. If the items do not comply with the specifications, then they are considered nonconforming. The traditional Shewhart np-chart is used to monitor the number of nonconforming units in a sample and is basically constructed from a 99.73% confidence interval. Due to technological advancement in manufacturing processes, more counts of zeros are observed. The excess of zeros under the binomial distribution is called Zero-Inflated Binomial (ZIB). If the np-chart is used to evaluate zero-inflated data, the control chart often underestimates the mean and variance of the zero-inflated counts, resulting in narrow control limits and subsequently leading to a higher false alarm rate in detecting out-of-control signals. The following research proposes remarkable solutions for this drawback.

Sim and Lim [1] introduced control charts for zero-inflated data in both a Poisson distribution and a binomial distribution. For the binomial distribution, four control charts were constructed to monitor the zero-inflated count: namely the traditional Shewhart (np-chart), a np-chart with ZIB model (np_{ZIB}-chart), a np-chart with Jeffreys prior interval (np_{J}-chart) and a np-chart with Blyth-Still interval (np_{BS}-chart). Using the ARL as an indicator, the study reveals that the np_{J}-chart and np_{BS}-chart with the two-of-two control rule perform better than the np-chart and the np_{ZIB}-chart in all situations.

Peerajit and Mayureesawan [2] carried out an extensive study of [2] for a Zero-Inflated Poisson (ZIP) distribution. The study compares the performance of a c-chart with a Jeffreys prior interval (c_{J}-chart), a c-chart with ZIP model (c_{ZIP}-chart) and a Shewhart c-chart (c-chart) in processes with different excess zero proportions, using the ARL and the ACP as performance indicators. The results reveal that the c_{J}–chart has higher efficiency than the other control charts when the process is in-control. When the process is out-of-control, the c-chart was the most effective control chart to detect the shift. However, the ACP of the c-chart was found to be too low to be a good estimator for the process parameters.

This paper compares the performance of control charts by applying the K of K control rule from the Sim and Lim [1] and uses the ACP concept from the Peerajit and Mayureesawan [2] study. The performance was compared in terms of the ARL and the ACP for different parameters, namely, the proportions of observed zeros (φ), variances (σ^2) and the shift in the nonconforming proportion (θ).

2. Materials and Methods

2.1 Control chart models

In order to compare the performance of the various control charts, all their relevant characteristics were collected from the literature used in this study.

1) np-chart

If p is a nonconforming fraction in a manufacturing process with a sample unit of n, the upper control limit (UCL) from the Shewhart method [3] is determined by

$$UCL = np + 3\sqrt{np(1-p)} \tag{1}$$

when p is estimated with $\bar{p}$.

2) np$_{ZIB}$-chart

Sim and Lim [1] proposed the ZIB model in a zero-Inflated binomial distribution. If Z is a ZIB random variable, the density function of Z is defined as

$$P(Z = z) = \phi I_{(z,0)} + (1-\phi)\, g(z;p)\,; \quad z = 0, 1, 2, ... \tag{2}$$

where $I_{(z,0)} = 1$ if $z = 0$ and $I_{(z,0)} = 0$ if z = others.

$g(z;p)$ is a binomial density function or

$$g(z;p) = C_z^n p^z (1-p).^{n-z}$$

The maximum likelihood estimates (MLEs) of the parameters p and ϕ are given as

$$\hat{p} = \frac{[1-(1-\hat{p})^n]\,\bar{z}^+}{n} \tag{3}$$

$$\hat{\phi} = 1-\frac{\bar{z}}{n\hat{p}} \tag{4}$$

where $\bar{z}^+$ is the mean of the m^+ positive observations z^+

$\bar{z}$ is the sample mean.

The estimation of p from (3) is also used in constructing the upper control limit of np$_{ZIB}$-chart and other type of control charts. The upper control limit of np$_{ZIB}$-chart is given as

$$UCL = n\hat{p} + 3\sqrt{n\hat{p}(1-\hat{p})}. \quad (5)$$

3) np_J-chart

Sim and Lim [1] suggested the npJ-chart using (1-α)100% Jeffreys prior interval which the upper control limit is given as

$$UCL = \max[x \mid p_0 \geq B(\alpha; x+0.5, n-x+0.5] \quad (6)$$

where X is a binomial random variable with parameters n and p and B(α; a, b) is the 100th percentile of a beta distribution with parameters a and b.

ρ_0 denotes the estimated value of p calculated from equation (3).

4) np_W-chart

The np_W-chart is constructed using the Wilson confidence interval. The interval was developed by Wilson [4]. This binomial interval is an improvement when the actual coverage probability is closer to the nominal value over the normal approximation interval. Let X be a binomial random variable with parameter n and p, the (1-α)100% confident interval of p is given as:

$$\frac{x + \frac{z_\alpha^2}{2}}{n + z_\alpha^2} \pm \frac{z_\alpha \sqrt{n}}{n + z_\alpha^2} \sqrt{\tilde{p}\tilde{q} + z_\alpha^2 / 4n}.$$

The upper control limit of the np_W-chart is given as

$$UCL = \max[x \mid p_0 \geq w(x)] \quad (7)$$

where $\tilde{p} = \frac{x}{n}$ and $w(x) = \frac{x + \frac{z_\alpha^2}{2}}{n + z_\alpha^2} - \frac{z_\alpha \sqrt{n}}{n + z_\alpha^2} \sqrt{\tilde{p}\tilde{q} + z_\alpha^2 / 4n}$.

p_0 denotes the estimated value of p calculated from equation (3).

5) np_{AC}-chart

The np_{AC}-chart is constructed by using the Agresti-Coull interval, an approximate binomial confidence interval as proposed by Agresti and Coull [5] and Brown et al. [6]. Given x successes in n trials, define

$$\tilde{p} = \tilde{x} / \tilde{n}, \quad \tilde{q} = 1 - \tilde{p}, \quad \tilde{x} = x + \frac{z_\alpha^2}{2}, \quad \tilde{n} = n + z_\alpha^2.$$

The confidence interval for p is given by

$$\tilde{p} \pm z_{\alpha}(\tilde{p}\tilde{q})^{1/2}\,\tilde{n}.^{-1/2}$$

The upper control limit of np_{AC}-chart is defined as

$$\text{UCL} = \max[x \mid p_0 \geq ac(x)] \qquad (8)$$

where $ac(x) = \tilde{p} - z_{\alpha}(\tilde{p}\tilde{q})^{1/2}\,\tilde{n}.^{-1/2}$

p_0 denotes the estimated value of p calculated from equation (3).

6) np_{BS}-chart

The np_{BS}-chart is suggested by Sim and Lim [1] for use with the Blyth-Still confidence interval [7]. Let X be a binomial random variable with parameter n and p. The upper control limit of this control chart is given as:

$$\text{UCL} = \max[x \mid p_0 \geq a(x)] \qquad (9)$$

where $a(x) = \dfrac{(x-0.5)+0.5z_{\alpha}^2 - z_{\alpha}\sqrt{(x-0.5)-\dfrac{(x-0.5)^2}{n}+0.25z_{\alpha}^2}}{n+z_{\alpha}^2}.$

p_0 denotes the estimated value of p calculated from equation (3).

2.2 Experimental design and methodology

To evaluate the performance of control charts, the Monte Carlo Simulation methodology is used. This study applies the K of K control rule to detect assignable causes if K consecutive points fall above the upper control limit. The values of K studied are 1, 2, 3, 4 and 5. The variance of generated data (σ^2) is defined at 2, 3, 4 and 5. The proportion of nonconforming (p_0) in this simulation is 0.004 and the shift of this proportion (θ) is studied for values 0, 0.2, 0.4 and 0.6. The proportion (φ) of observed zeros is studied in 0.3(0.1)0.9. The experiment is based on 5,000 samples, repeated 50,000 times for each case. The performance evaluation of the control charts, based on the Average Run Length (ARL) and the Average Coverage Probability (ACP) is computed with the following criteria.

1. ARL: for the in-control situation, the effective charting method is the one with the in-control average run length (ARL_0) close the desirable value of 370. For the out of

control case, the preferred control chart is the one with low out-of-control average run length (ARL_1).

2. ACP: for a given process situation, the control chart that generates the ACP value close to the confidence coefficient (1-α) is the recommended control chart for that situation.

3. Results

The ARL values and the ACP values for the control charts from simulations were found as follows. The quantity ARL_0-DIFF is defined as the absolute value of the difference between ARL_0 and 370, the ACP-DIFF is the absolute value of the difference between ACP and 1-α.

Below is a summary of the results of the performance comparisons for some values of K, σ^2, φ and θ for the case where the performance is in-control and out-of-control.

3.1 The in-control performance

Table 1 shows the ARL_0 values and the ACP values for K = 1, 3, 5, σ^2 = 2, 5 and φ = 0.3, 0.5, 0.7, 0.9. Figure 1 and Figure 2 present the ARL_0-DIFF and the ACP-DIFF respectively when K = 1, for all values of σ^2. In this case, the np_j-chart is superior to others for low φ (0.3-0.4) and the np_w-chart outperforms the others for high φ (0.9) in all values of σ^2. Figure 3 and Figure 4 show the ARL_0-DIFF and the ACP-DIFF respectively when K = 3. These figures reveal that for lower φ (0.3-0.6) the np_j-chart is the preferred choice, but for higher φ (0.8-0.9) the np-chart is more desirable for all levels of σ^2. Figure 5 and Figure 6 present the ARL_0-DIFF and the ACP-DIFF respectively when K = 5. In this case we found that the np_j-chart is more preferable if φ is low (0.3) and the np-chart has a better performance than the others when φ is high (0.6-0.9) for all levels of σ^2.

Table 2 shows the preferable control charts for the in-control situation for each value of K, σ^2 and φ.

Table 1. The comparison of ARL_0 and ACP when $\theta = 0$.

K	σ^2	φ	ARL_0						ACP					
			NP	ZIB	JF	AC	WS	BS	NP	ZIB	JF	AC	WS	BS
1	2	0.3	26	84	250	250	85	84	0.9628	0.9917	0.9971	0.9994	0.9917	0.9967
		0.5	12	117	299	298	118	117	0.9370	0.9939	0.9979	0.9993	0.9936	0.9976
		0.7	9	183	363	361	182	183	0.9046	0.9963	0.9990	0.9997	0.9961	0.9990
		0.9	15	338	446	445	339	340	0.9415	0.9989	0.9995	0.9999	0.9987	0.9993
	5	0.3	19	221	221	221	101	220	0.9519	0.9953	0.9981	0.9983	0.9955	0.9963
		0.5	7	271	271	270	138	270	0.8965	0.9966	0.9985	0.9986	0.9967	0.9976
		0.7	4	339	339	340	210	339	0.8526	0.9977	0.9993	0.9992	0.9979	0.9983
		0.9	10	436	438	436	358	437	0.9114	0.9993	0.9996	0.9998	0.9991	0.9994
3	2	0.3	105	400	401	106	400	401	0.7703	0.9596	0.8998	0.8998	0.8995	0.9309
		0.5	230	459	459	232	459	459	0.7987	0.9722	0.9273	0.9273	0.9289	0.9519
		0.7	151	490	490	406	491	490	0.8229	0.9821	0.9580	0.9580	0.9569	0.9701
		0.9	432	499	495	495	499	499	0.9131	0.9931	0.9848	0.9848	0.9858	0.9899
	5	0.3	66	416	285	216	280	415	0.7296	0.9503	0.9079	0.9079	0.9066	0.9116
		0.5	59	465	349	349	349	466	0.7052	0.9670	0.9320	0.9320	0.9318	0.9395
		0.7	72	492	458	459	459	492	0.7420	0.9810	0.9588	0.9588	0.9569	0.9641
		0.9	412	499	498	498	498	499	0.9025	0.9943	0.9856	0.9856	0.9847	0.9884
5	2	0.3	131	446	446	445	446	446	0.6138	0.7710	0.7745	0.7734	0.7707	0.8825
		0.5	341	488	488	488	488	488	0.7071	0.8373	0.8350	0.8380	0.8376	0.9120
		0.7	405	499	499	499	499	499	0.7595	0.9039	0.9037	0.9025	0.9024	0.9444
		0.9	499	500	500	500	500	500	0.9143	0.9691	0.9653	0.9692	0.9696	0.9796
	5	0.3	162	388	388	388	388	486	0.6034	0.7775	0.7413	0.7760	0.7795	0.8328
		0.5	204	473	475	474	474	497	0.6243	0.8415	0.8164	0.8410	0.8414	0.8816
		0.7	400	498	498	498	498	498	0.7288	0.9051	0.8941	0.9075	0.9052	0.9260
		0.9	499	500	500	500	500	500	0.9001	0.9683	0.9653	0.9679	0.9685	0.9761

Note : NP = np-chart, ZIB = np_{ZIB}-chart, J F = np_j-chart, AC = np_{AC}-chart, WS = np_w-chart, BS = np_{BS}-chart.

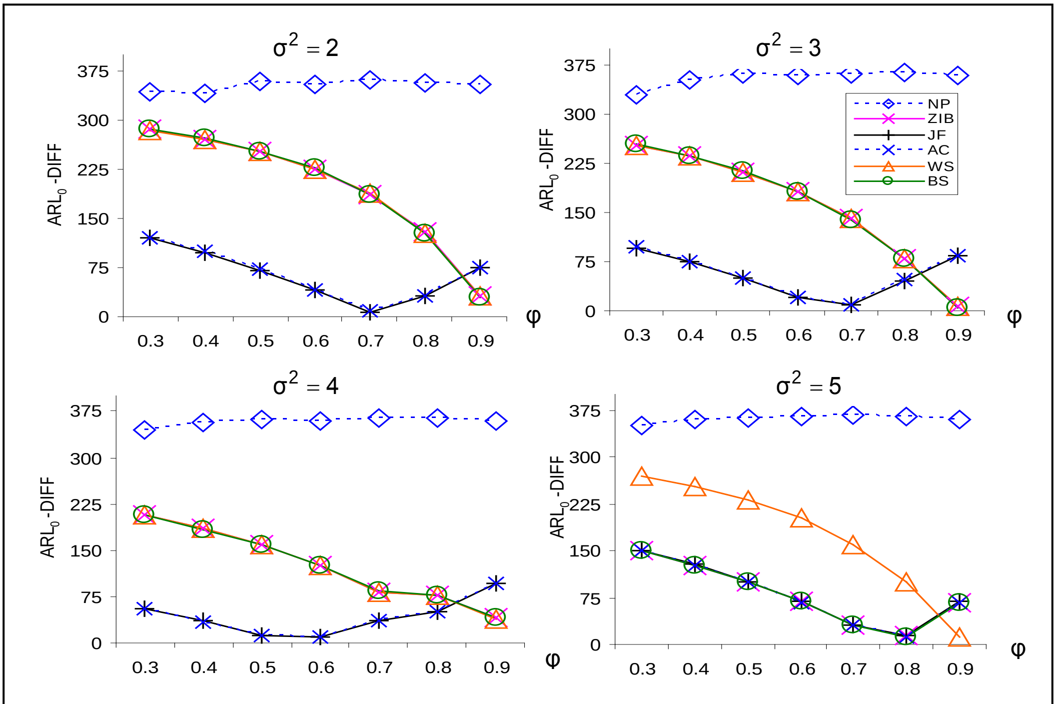

Figure 1. The comparison of the ARL_0 – DIFF for K = 1 and θ = 0.

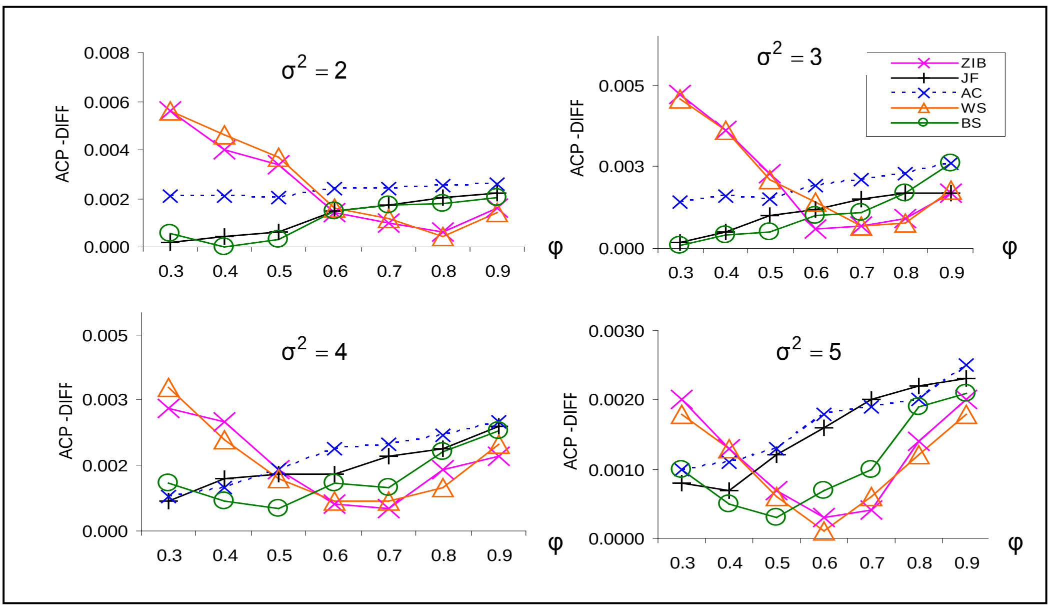

Figure 2. The comparison of the ACP- DIFF for K = 1 and θ = 0.

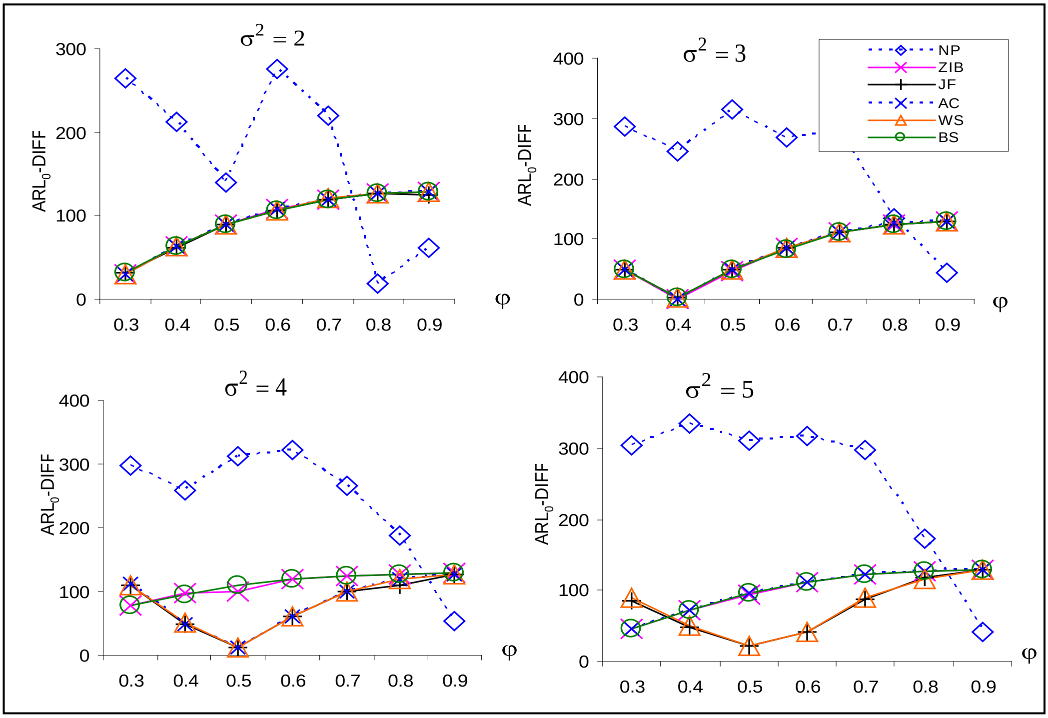

Figure 3. The comparison of the ARL$_0$- DIFF for K = 3 and θ = 0.

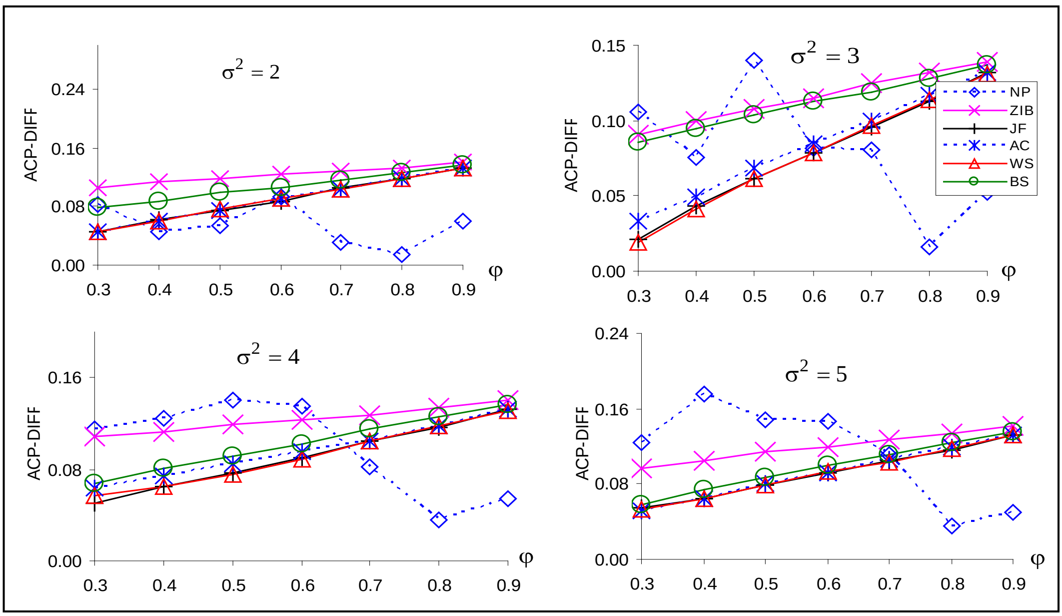

Figure 4. The comparison of the ACP- DIFF for K = 3 and θ = 0.

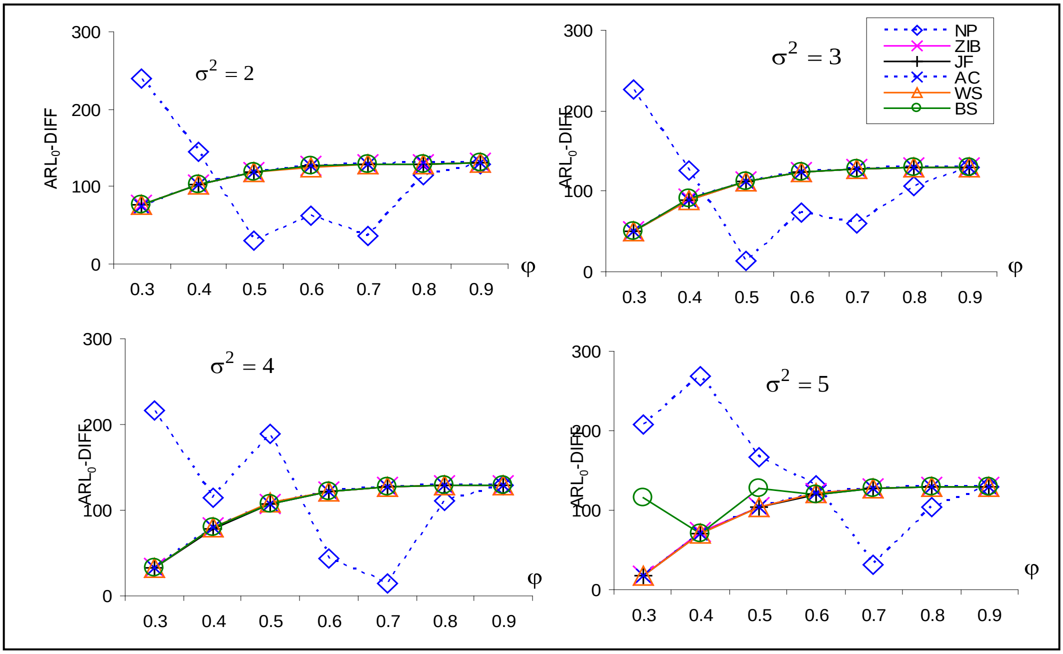

Figure 5. The comparison of the ARL_0- DIFF for K = 5 and θ = 0.

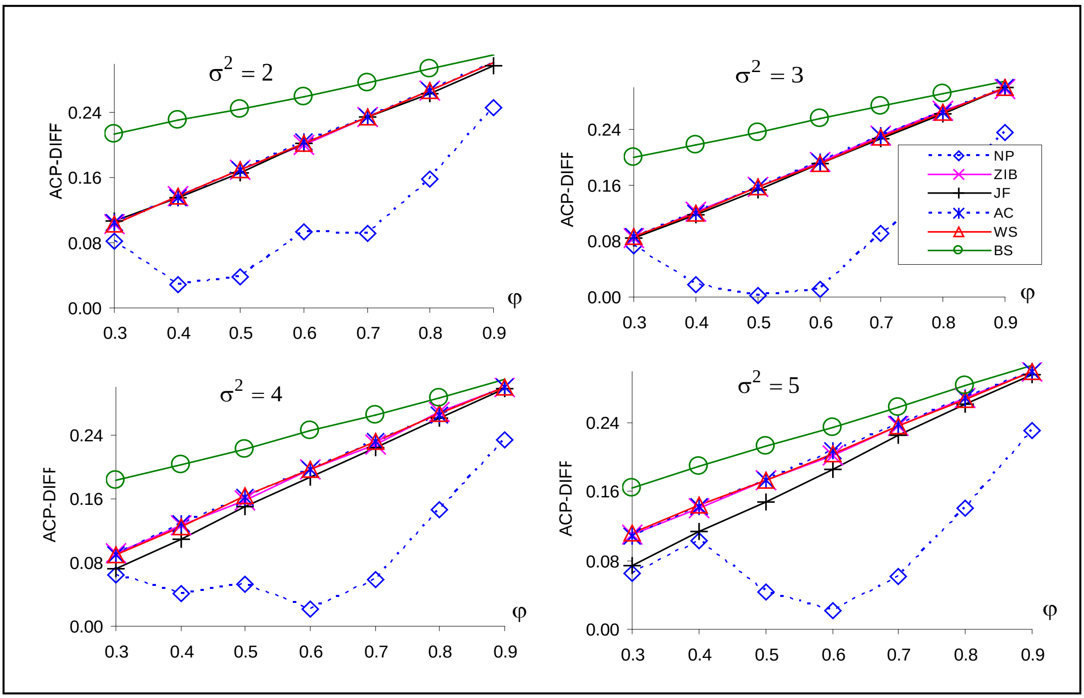

Figure 6. The comparison of the ACP- DIFF for K = 5 and θ = 0.

Table 2. The preferable control charts for the in-control situation.

K	σ^2	φ	The preferable control charts
1	2 and 3	0.3 -0.6	np_j-chart
		0.7-0.8	-
		0.9	np_W-chart
	4 and 5	0.3-0.4	np_j-chart
		0.5-0.8	-
		0.9	np_W-chart
2	All values	0.3 -0.6	np_{ZIB}-chart
		0.7-0.9	np_W-chart
3	2	0.3 -0.6	np_j-chart
		0.7	-
		0.8-0.9	np-chart
	3, 4 and 5	0.3-0.7	np_j-chart
		0.8-0.9	np-chart
4	2	0.3 -0.5	np_{ZIB}-chart
		0.6-0.9	np-chart
	3, 4 and 5	0.3-0.4	np_{ZIB}-chart
		0.5-0.6	-
		07-0.9	np-chart
5	All values	0.3	np_j-chart
		0.4-0.6	-
		0.7-0.9	np-chart

3.2 The out-of-control performance

In Figure 7 and Figure 8 we compare the ARL_1 and the ACP-DIFF when K = 3, σ^2 = 2, 5 and θ = 0.2, 0.6. For σ^2 = 2, it clearly shows that the np_j-chart is preferred for low φ (0.3-0.5) and the np-chart is preferred for high φ (0.8-0.9) in all levels of θ. Similarly, for σ^2 = 5 case, the np-chart is found to be a more appropriate control chart when φ is high (0.8-0.9) for all levels of θ.

In Table 3 we show the preferable control charts for the out-of-control situation for each value of K, σ,^{2} φ and θ.

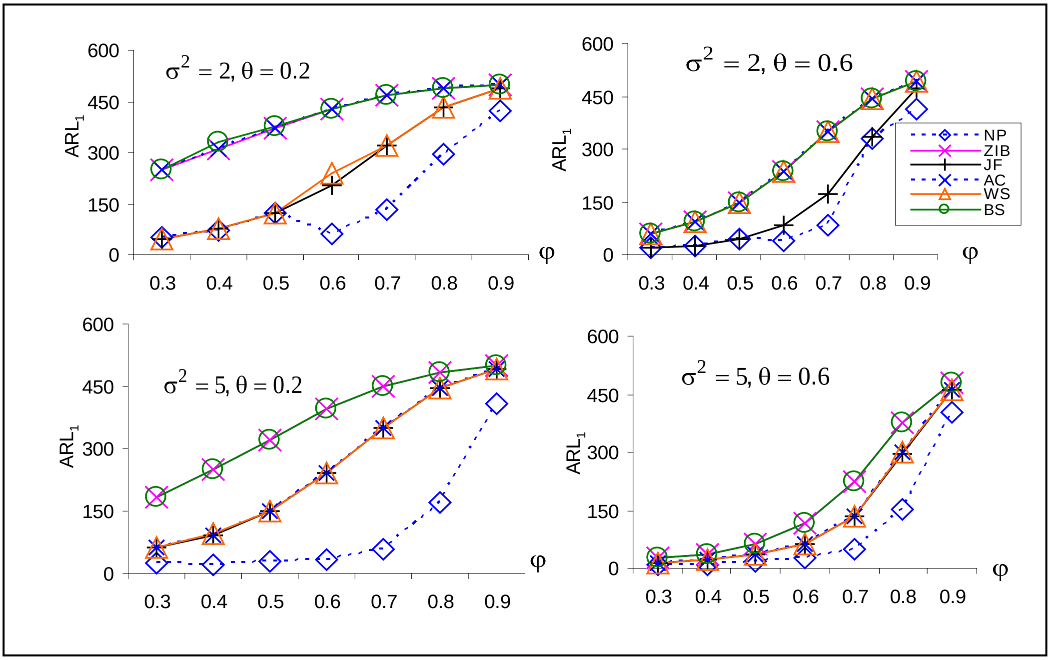

Figure 7. The comparison of ARL_1 when K =3 and, θ≠0.

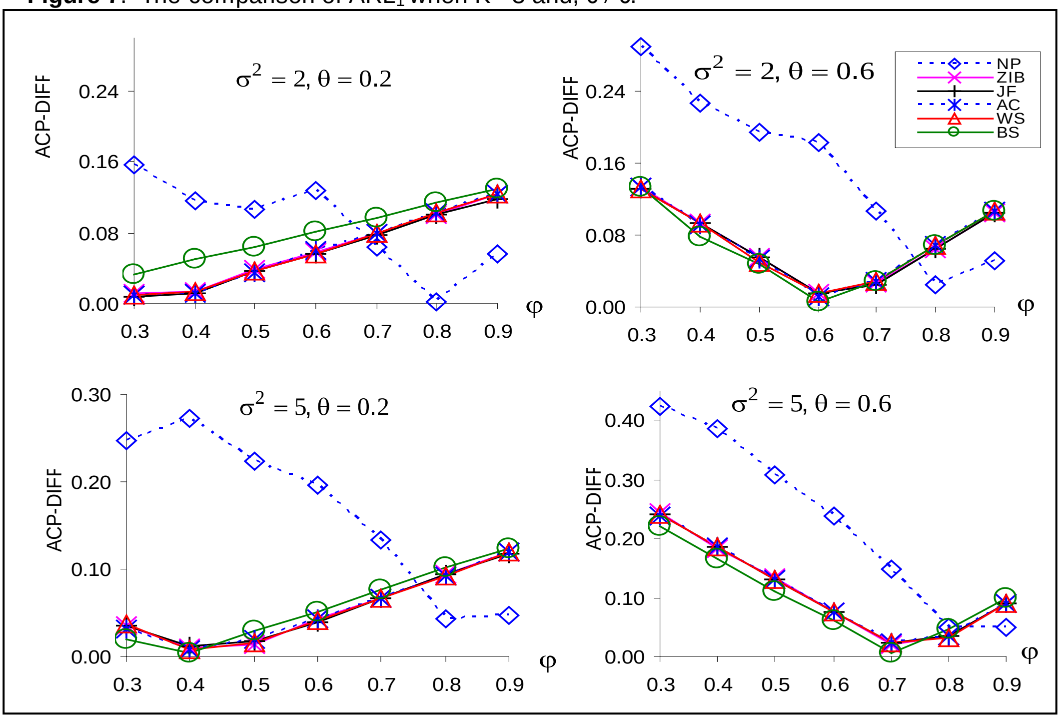

Figure 8. The comparison of ACP - DIFF when K = 3 and θ ≠ 0.

Table 3. The preferable control charts for the out-of -control situation.

K	σ^2	θ	φ	The preferable control charts
1	2, 3 ,4 and 5	All values	0.3-0.9	-
2	2		0.3	np_j-chart
			0.4-0.8	-
			0.9	np-chart
	3, 4 and 5		0.3-0.8	-
			0.9	np-chart
3	2		0.3-0.5	np_J-chart
			0.6-0.7	-
			0.8-0.9	np-chart
	3, 4 and 5		0.3-0.7	-
			0.8-0.9	np-chart
4	2 and 3		0.3	np_J-chart
			0.4-0.6	-
			0.7-0.9	np-chart
	4 and 5		0.3-0.6	-
			0.7-0.9	np-chart
5	2		0.3-.4	-
			0.5-0.9	np-chart
	3, 4 and 5		0.3-0.5	-
			0.6-0.9	np-chart

4. Discussion and Conclusions

In this paper, we compare the performance of the control charts based on the binomial distribution when the occurrence of zero inflated counts is observed. The control charts are based on different binomial confidence interval methods. The average run length and the average coverage probability are used to measure the performance of the control charts for different values of variances, the observed zero proportions and the shift in nonconforming proportions. The control chart signals the out-of-control process when K consecutive observation exceeds the upper control limit, the K of K control rule.

The results show that for the in-control case, the np_j-chart performs better than the other control charts when K = 1, 3 and 5 and the zero proportion is relatively low (φ = 0.3) in all levels of variance. On the other hand, the np_{ZIB}-chart seems to perform better than the others when K = 2 and 4 and the zero proportion is low (φ = 0.3 and 0.4) in all levels of variance. It should be noted that the np-chart is more suitable for high K (K = 3, 4, 5) and high observed zero proportion (φ = 0.8-0.9) in all levels of variance. For the out-of-control case, when K is 2, 3, 4, the np_J-chart has higher efficiency than the other control charts for low variance and low zero proportion (σ^2 = 2, φ = 0.3) in all levels of shift. In addition, the np-chart is more fitting when K = 2, 3, 4, 5 and zero proportion is considerably high (φ = 0.9) in all levels of S^2 and all magnitudes of shift in nonconforming proportion.

It can be concluded that in situations when the low zero proportion (φ =0.3) and the low variance (σ^2 = 2) are observed, the np_j-chart is recommended for use with applying the K of K control rule when K is 2 or higher. However, if the zero proportion is extremely high (0.9), the np-chart is suggested when applying the K of K control rule if K is 3 or higher. It should be noted that for moderate to high zero proportions (0.4-0.8), the studied control charts did not perform well. Therefore, this problem could be an interesting issue for further study.

5. Acknowledgements

The authors would like to thank the Graduate College, King Mongkut's University of Technology North Bangkok for the financial support during this study.

References

[1] Sim, C.H., and Lim, M.H., Attribute Charts for Zero-Inflated Processes. Commun Stat Simut C, 2008; 37: 1440-1452.

[2] Peerajit V., and Mayureesawan T., Nonconforming Control Charts for Zero-Inflated Processes, Proceedings of the National Conference on Statistics and Applied Statistics, 2010; 61-73.

[3] Montgomery, D.C., Introduction to Statistical Quality Control. 4th Edition, New York, McGraw-Hill, 2005.

[4] Wilson, E.B., Probable inference, the law of succession, and statistical inference. JASA, 1927; 22: 209-212.

[5] Agresti, A., and Coull, B.A., Approximate is better than exact for interval estimation of binomial proportions. American Statistician, 1998; 52: 119-126.

[6] Brown, L.D., Cai, T., and DasGupta, A., Interval estimation for a binomial proportion (with discussion). Statistical Science, 2001; 16: 101-133.

[7] Blyth, C.R., and Still, H.A., Binomial confidence intervals. JASA, 1983; 78: 108-116.