CONTROL CHARTS FOR BINOMIAL WHEN THE UNDERLYING DISTRIBUTION IS RATIO OF TWO POISSON MEANS

Ashit B. Chakraborty * and Anwer Khurshid**[1]
*St. Anthony's College, Shillong, Meghalaya, Indi
**Department of Mathematical and Physical Sciences, College of Arts and Science, University of Nizwa, P. O. Box 33, PC 616, Birkat Al Mouz, Oman

ABSTRACT:
In this paper, an attempt has been made to construct control charts for binomial distribution when the underlying distribution is ratio of two Poisson means. Shewhart chart control limits and ARL are obtained based on exact value of the confidence limit.

KEY WORDS: binomial confidence interval; control chart; average run length; Poisson distribution

MSC: 62F03

RESUMEN
En este trabajo se ha hecho un intento de construir cartas de control para la distribución binomial cuando la distribución subyacente es la razón de dos medias Poisson. Limites de cartas de control de Shewhart y ARL son obtenidos en base al valor exacto del límite de confianza.

1. INTRODUCTION

A common problem in the application of statistical methods to the quality of material produced by a continuous process is to ensure that the proportion of defective product does not exceed a specified limit. In 1924 Walter A. Shewhart developed the most important statistical tool in statistical process control: the control chart. Traditionally Shewhart control chart assumes normality. Alternatively to control charts based on normality, control charts based on other parametric distributions have been proposed. For example, Ferrel (1958) and Cheng and Xie (2000) proposed a control chart based on lognormal distribution; Nelson (1979) obtained control limits for median, range and location scales for the Weibull distribution; Kaminsky et al. (1992) developed control charts based on geometric distribution. Xie and Goh (1997) and Schwertman (2005) illustrate the use of geometric and negative binomial distributions for constructing control charts.

Shewhart control chart based on Poisson distribution is used when the response from a process is a count such as, number of accidents per worker in a factory, number of persons suffering from an infectious disease, the number of sibs with human albinism in families of any size. Literature is also available on control charts based on counted data. Lucas (1985) described the design and implementation procedure for counted data for detection of increase or decrease in the count level. Chakraborty and Singh (1990) constructed Shewhart control charts for zero-truncated Poisson distribution where average run length (ARL) and operating characteristic (OC) function were obtained. Recently Chen et al. (2008) obtained attribute control charts using generalized zero-inflated Poisson distribution.

However, in many situations the traditional technique of Shewhart control charts may not be suitable or can not be used, as for many processes, the assumptions of Poisson distribution may deviate or may provide inadequate model. Distribution of counts generated by various types of processes can not be modeled by the Poisson distribution to use in for c-chart. Recently, Hoffman (2003) developed control limits based on negative binomial for counted data with extra Poisson variation.

[1] anwer_khurshid@yahoo.com,
**Corresponding author

The problem of comparing two Poisson rates has been studied in the literature, for example, Nelson (1987), Sahai and Misra (1992), Sahai and Khurshid (1993), Price and Bonett (2000), Bratcher and Stamey (2004), Krishnamoorthy and Thompson (2004), and Gu et al. (2008) to mention a few. In many cases, binomial distribution may be more flexible and natural to use in place of Poisson distribution, when the control charts for ratio of two Poisson means (Sahai and Khurshid, 1993) need to be constructed, as the situation may arise to control the ratio rather than the single parameter. In such cases traditional 3σ $(k\sigma)$ chart may not be ideal to detect the shift of the ratio of two means and since binomial is an asymmetric distribution (for $p \neq q$), the probability (exact) limits are more appropriate than typical 3σ control limits, as Shewhart control charts are based on normal distribution.

This paper considers the binomial model as a simple and flexible alternative to the Poisson model (which may be the case of over-dispersion also) for detecting the shift of the ratios of process parameters of two Poisson distributions. Control limits of Shewart control charts are obtained for binomial distribution when the underlying distribution is Poisson and hence the limits of the control charts are obtained based on the upper limit of the confidence intervals for the ratio of two Poisson means. In most of the cases for attribute control charts, the lower control limits may not exist because the probability at zero could be larger than the desired type I error probability, hence in this paper, we have considered control limits based on the value of exact upper confidence limit. ARL of the chart is studied for different values of the parameters of the distribution for different control limits.

2. BINOMIAL DISTRIBUTION (WHEN UNDERLYING DISTRIBUTION IS POISSON)

Let X and Y be two independent Poisson variates with parameters λ and μ respectively. Then the conditional distribution of X given $X + Y$ follows a binomial distribution with parameters n and $p = \dfrac{\lambda}{\lambda+\mu}$ (Lehmann and Romano, 2005)

$$P\left[X \middle| (X+Y=n)\right] = P(X = x \cap Y = n-x)\left[P(X+Y=n)\right]^{-1}$$
$$= \binom{n}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^{x}\left(\frac{\mu}{\lambda+\mu}\right)^{n-x} \qquad (2.1)$$

where $x = 0,1,2,...,n$. The distribution (2.1) is used to obtain control limits for Shewhart chart based on exact method for $100(1-\alpha)\%$ confidence interval on $\theta = \dfrac{\lambda}{\mu}$.

3. CONTROL CHARTS AND CONSTRUCTION OF CONTROL LIMITS

3.1 Shewhart ($k\sigma$) control charts

For the above distribution (2.1), the mean and variance are given by:

$$E(X) = \frac{n\lambda}{\lambda+\mu}, \; Var(X) = \frac{n\mu\lambda}{(\lambda+\mu)^2}.$$

The limits for np and fraction defective charts are as follows:

$$\frac{n\lambda}{(\lambda+\mu)} \pm K\sqrt{\frac{n\lambda\mu}{(\lambda+\mu)^2}}$$

and

$$\frac{\lambda}{(\lambda+\mu)} \pm K\sqrt{\frac{\lambda\mu}{n(\lambda+\mu)^2}}.$$

where K is standardized normal variate.

3.2 Exact Control Limits

For a specified level α (Type I error), upper and lower control limits are $P(x > UCL) \le \alpha$ and $P(x < UCL) \le \alpha$, i.e.,

$$P(X \le x | p = p_U) = \sum_{k=0}^{n} \binom{n}{k} p_U^k (1 - p_U)^{n-k} \le \frac{1}{2}\alpha$$

and

$$P(X \ge x | p = p_L) = \sum_{k=x}^{n} \binom{n}{k} p_L^k (1 - p_L)^{n-k} \le \frac{1}{2}\alpha .$$

Hence for exact α -level control limits, we have UCL such that

$$\sum_{UCL}^{n} \binom{n}{k} p_U^k (1 - p_U)^{n-k} > \frac{1}{2}\alpha$$

and

$$\sum_{UCL+1}^{n} \binom{n}{k} p_U^k (1 - p_U)^{n-k} < \frac{1}{2}\alpha$$

where p_U is the upper confidence limits, when the limits are set on $p = \frac{\lambda}{\lambda + \mu}$ and here $p_U = 1 - \left(\frac{\alpha}{2}\right)^{1/n}$.

4. AVERAGE RUN LENGTH (ARL)

ARL is the average number of points that must be plotted before a point indicates an out of control condition. For any Shewhart control chart, the ARL is defined as $ARL = [P]^{-1}$, where P is the probability that a single point exceeds the control limits. Now, if the mean shifts from the in-control value, say, θ_0, to another value $\theta_1 = \theta_0 + k\sigma$, the probability of not deleting this shift on the first subsequent sample or the β -risk (Montgomery, 2005) is

$$\beta = P\{x < UCL | \theta\} - P\{x \le LCL | \theta\}.$$

Thus for (2.1), we have

$$\beta = \sum_{x=0}^{UCL} \binom{n}{x} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right)^{n-x} - \sum_{x=0}^{LCL} \binom{n}{x} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right)^{n-x} \tag{4.1}$$

Hence,

$$\begin{aligned} P &= 1 - \beta \\ &= 1 - \left[\sum_{x=0}^{UCL} \binom{n}{x} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right)^{n-x} - \sum_{x=0}^{LCL} \binom{n}{x} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right)^{n-x} \right] \end{aligned}$$

which will give ARL as

$$ARL = \left[1 - \sum_{x=0}^{UCL}\binom{n}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^{x}\left(\frac{\mu}{\lambda+\mu}\right)^{n-x} + \sum_{x=0}^{LCL}\binom{n}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^{x}\left(\frac{\mu}{\lambda+\mu}\right)^{n-x}\right]^{-1} \tag{4.2}$$

Operating Characteristic (OC) curve for the Shewhart control chart when the underlying distribution is binomial can be constructed by plotting the β-risk against the magnitude of the shift of the (process) parameter that we wish to detect.

5. AN ILLUSTRATIVE EXAMPLE AND CONCLUSION

An example of counts of diatoms (Sahai and Khurshid, 1993) is considered to develop control limits, where 95% confidence interval on $\theta = \frac{\lambda}{\mu}$ was constructed when the sampling distribution is restricted to $n = 25$ counts. λ and μ being the mean diatom concentration of two lake waters.

The exact upper confidence limit (θ) from the example of diatom is 2.1990 which is calculated from $p = \frac{\lambda}{\lambda+\mu} = 0.6874$. This upper limit of $p(=p_U)$ is used in this paper to develop the limits of control charts and hence average run length (ARL).

The 3σ Shewhart control chart limits for $p = 0.6874$ are $UCL = 24$ and $LCL = 10$. ARL values are calculated for change in the values of p. ARL also calculated for different control limits as shown in Tables 1 and 2.

Table 1: **Some numerical values of β for $p_0 = 0.6847 \cong 0.7$ and different control limits**

	β				
p_0	$(\mu \pm 3\sigma)$ $UCL = 24,$ $LCL = 10$	$(\mu \pm 2\sigma)$ $UCL = 22,$ $LCL = 13$	$(\mu \pm 1.5\sigma)$ $UCL = 21,$ $LCL = 14$	$(\mu \pm \sigma)$ $UCL = 20,$ $LCL = 15$	$(\mu \pm 0.5\sigma)$ $UCL = 18,$ $LCL = 16$
0.2	0.006	0	0	0	0
0.3	0.098	0.002	0	0	0
0.4	0.414	0.035	0.007	0.001	0
0.5	0.788	0.202	0.067	0.013	0
0.6	0.966	0.556	0.029	0.096	0.005
0.7	0.998	0.880	0.671	0	0.045
0.8	0.996	0.985	0.937	0.362	0.223
0.9	0.928	0.911	0.901	0.877	0.57
0.95	0.723	0.693	0.676	0.659	0.556

The values of β and ARL for shifting of the parameter $p_0(\cong 0.6874)$ to some other values p_1 (say) are shown in Tables 1 and 2 for different control limits. It is evident from the Table 1 that the values of β (the probability of not detecting the shift from p_0 to p_1 on the first subsequent sample) will go on increasing for fixed control limits $\mu \pm k\sigma$ $(k = 3, 2, 1.5, 1, 0.5)$ as we keep increasing the shift of the parameter from p_0 to p_1 but for fixed shift of the parameter, the values of β decrease as we increase the size of the control limit. Whereas from the Table 2 it is observed that the values of ARL for fixed control limits $\mu \pm k\sigma$ $(k = 3, 2, 1.5, 1, 0.5)$ will go on decreasing as we keep increasing the shift of the parameter from p_0 to p_1. But for fixed shifting of the parameter, the values of ARL increase as we increase the size of the control limit.

Table 2: **Some numerical values of ARL for** $p_0 = 0.6847 \cong 0.7$ **and different control limits**

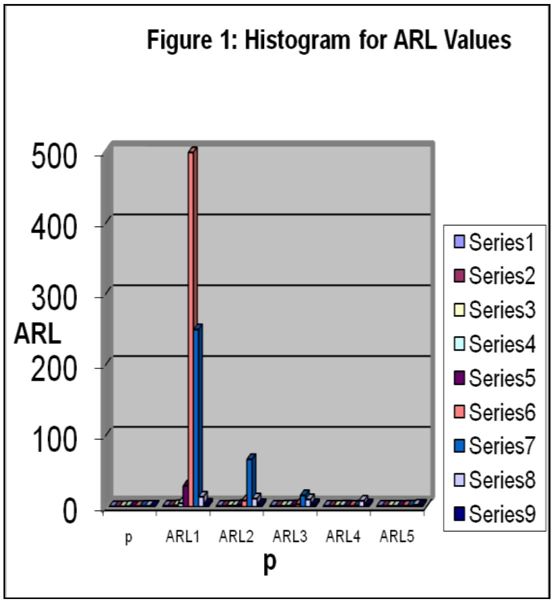
p_0	ARL_1 $(\mu \pm 3\sigma)$	ARL_2 $(\mu \pm 2\sigma)$	ARL_3 $(\mu \pm 1.5\sigma)$	ARL_4 $(\mu \pm \sigma)$	ARL_5 $(\mu \pm 0.5\sigma)$
0.2	1.006	1	1	1	1
0.3	1.1086	1.002	1	1	1
0.4	1.7064	1.0363	1.007	1.001	1
0.5	4.7169	1.253	1.072	1.0132	1
0.6	29.412	2.252	1.408	1.106	1.005
0.7	**500**	8.333	3.0395	1	1.047
0.8	280	66.667	15.874	1.567	1.287
0.9	13.88	11.2359	10.1	8.13	2.326
0.95	3.61	3.257	3.086	2.932	2.252

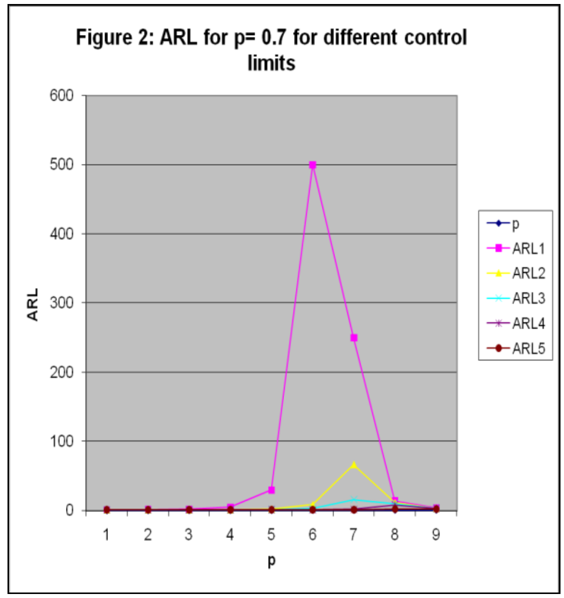
Table 3 shows the values of ARL for different values of n for fixed $p = 0.6874 \cong 0.7$. It has been observed here that for fixed n, the values of ARL decrease as we decrease the range of the limit, whereas ARL values are random in nature as n takes different values for fixed limits. This interesting case will be considered in detail elsewhere.

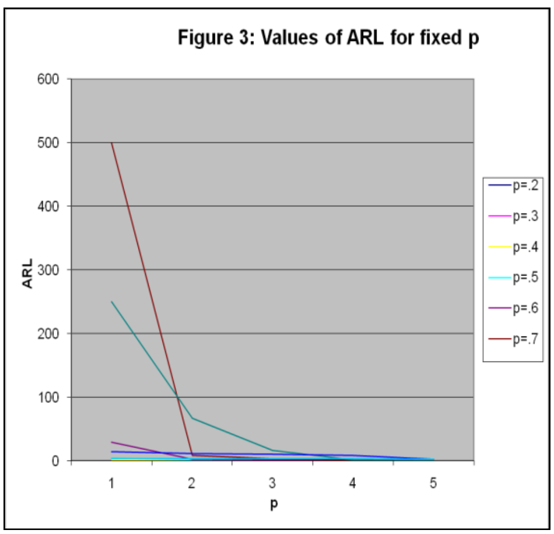
Table 3: **Values of ARL for different control limits and** n **when** $p = 0.7$

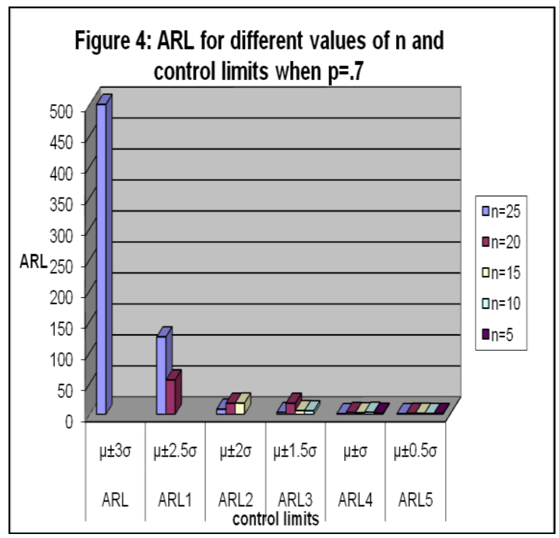
	n				
Limits	25	20	15	10	5
$(\mu \pm 3\sigma)$	500	-	-	-	-
$(\mu \pm 2\sigma)$	125	55.36	-	-	-
$(\mu \pm 2\sigma)$	8.3	17.86	18.18	-	-
$(\mu \pm 1.5\sigma)$	3.04	17.86	6.02	5.62	-
$(\mu \pm \sigma)$	1	2.99	2.47	3.34	1.76
$(\mu \pm 0.5\sigma)$	1.05	1.59	1.74	1.36	1.56

Figure 1 gives the idea of the distribution of ARL. The histogram shows the high degree of skewness in the positive direction as we decrease σ-limits and for a specified α, the test will detect large differences more easily if the σ-limits are decreased which is also observed in Figures 2 and 3. Figure 4 depicts graphic representation of ARL for different values of n and subsequent control limits for fixed $p = 0.6874 \cong 0.7$.









RECEIVED MARCH 2011
REVISED SEPTEMBER 2011

REFERENCES

[1] BRATCHER, T. L. and STAMEY, J. D. (2004): A note on Bayesian interval estimators for comparing two Poisson rates. **The Mathematical Scientist**, 29, 54-60.

[2] CHAKRABORTY, A. B. and SINGH, B. P. (1990): Shewhart control chart for zero truncated Poisson distribution. **Proceeding of National Seminar on Quality Reliability**, 18-24, Organized by National Institute for Quality and Reliability, Trivandrum Branch, India.

[3] CHEN, N., ZHOU, S., CHANG, T-S. and HUANG, H. (2008): Attribute control charts using generalized zero-inflated Poisson distribution. **Quality and Reliability Engineering International**, 24, 793-806.

[4] CHENG, S. W. and XIE, H. (2000): Control charts for lognormal data. **Tamkang Journal of Science and Technology,** 3, 131-137.

[5] FERREL, E. D. (1958): Control charts for lognormal universe. **Industrial Quality Control**, 15, 4-6.

[6] GU, X., NG, H. K. T., TANG, M. L. and SCHUCANY, W. R. (2008): Testing the ratio of two Poisson rates. **Biometrical Journal**, 50, 283-298.

[7] HOFFMAN, D. (2003): Negative binomial control limits for counted data with extra-Poisson variation. **Pharmaceutical Statistics**, 2, 127-132.

[8] KAMINSKY, F. C., BENNEYAN, J. C., DAVIS, R. D. and BURKE, R. J. (1992): Statistical control charts based on a geometric distribution**. Journal of Quality Technology**, 24, 63- 69.

[9] KRISHNAMOORTHY, K. and THOMSON, J. (2004): A more powerful test for comparing two Poisson means. **Journal of Statistical Planning and Inference**, 119, 23-35.

[10] LEHMANN, E. L. and ROMANO, J. P. (2005): **Testing Statistical Hypotheses**, Third Edition. Springer, New York.

[11] LUCAS, J. M. (1985): Counted data CUSUM's. **Technometrics,** 27, 29-44.

[12] MONTGOMERY, D. C. (2005): **Introduction to Statistical Quality Control**, Fifth Edition, John Wiley, New York.

[13] NELSON, L. S. (1987): Comparison of Poisson means: The general Case. **Journal of Quality Technology**, 19,173-179.

[14] NELSON, P. R. (1979): Control charts for Weibull processes with standards given. **IEEE Transactions on Reliability**, 28, 383-387.

[15] PRICE, R. M. and BONETT, D. G. (2000): Estimating the ratio of two Poisson rates. **Computational Statistics and Data Analysis**, 34, 345-356.

[16] SAHAI, H. and KHURSHID, A. (1993): Confidence intervals for ratio of two Poisson means. **The Mathematical Scientist,** 18, 43-50.

[17] SAHAI, H. and MISRA, S. C. (1992): Comparing means of two Poisson distributions. **The Mathematical Scientist,** 17, 60-67.

[18] SCHWERTMAN, N. C. (2005): Designing accurate control charts based on the geometric and negative binomial distributions. **Quality and Reliability Engineering International**, 21, 743-756.

[19] XIE, M. and GOH, T. N. (1997): The use of probability limits for process control based on geometric distribution. **International Journal of Quality and Reliability Management**, 16, 64-73.