

CUSUM CONTROL CHARTS FOR ZERO-TRUNCATED NEGATIVE BINOMIAL AND GEOMETRIC DISTRIBUTIONS

Anwer Khurshid*¹ and Ashit B. Chakraborty**²

*Department of Mathematical and Physical Sciences, College of Arts and Sciences, University of Nizwa, P. O. Box 33, PC 616, Birkat Al Mouz, Oman

**St. Anthony's College, Shillong, Meghalaya, India

ABSTRACT

In this paper CUSUM control charts for zero-truncated negative binomial distribution (ZTNBD) and zero-truncated geometric distribution (ZTGD) are constructed. Average run length (ARL) is studied for different values of the parameters of both the distributions. The method of Johnson (1961) is used for constructing the CUSUM chart.

KEYWORDS: control chart, Average Run Length (ARL), zero truncated negative binomial distribution (ZTNBD), zero-truncated geometric distribution (ZTGD):

MSC: 62P30

RESUMEN

En este trabajo se construyen gráficos de control del tipo CUSUM para la distribución binomial negativa cero-truncada (ZTNBD) y para la cero-truncada distribución geométrica (ZTGD): El largo de la racha promedio (ARL) es estudiada para diferentes valores de los parámetros de ambas distribuciones. El método de Johnson (1961) es utilizado para construir los gráficos CUSUM.

1. INTRODUCTION

Poisson distribution plays an important role in statistical quality control process through modeling random counts or control of defects per unit. Various types of processes can generate distributions of counts which can be modeled suitably by distributions different than Poisson distribution. Such processes include situations where counts tend to occur in clusters or where the intensity rate of the counts varies randomly over time. The negative binomial distribution (NBD) is a natural and more flexible extension of the Poisson distribution which allows for over-dispersion compared to the Poisson distribution (Hoffman, 2003). The application of NBD has been demonstrated in accident statistics, econometrics, quality control and biometrics. It is a well-known fact that geometric distribution (GD) is a special form of NBD. For detailed description, refer to Johnson et al. (2005), Khurshid et al. (2005) and Ryan (2011) among others.

Cumulative sum (CUSUM) control charts are widely used monitoring processes with the objective of improving process quality and productivity (Luceno and Puig-Pey, 2000). The pioneering work on CUSUM control charts is attributed to Page (1954). Lucas (1985) described the design and implementation procedure for count data (accidents) through CUSUM chart to detect increase or decrease in the count level. Several different types of control charts based on Poisson distribution are available in the literature. When the objective of any sampling plan of statistical quality control is to continue sampling until a certain number of successes has been achieved, then the number of items sampled will follow a NBD. Unfortunately, the literature on the control charts for the NBD and GD is scanty (Kaminsky et al., 1992; Ma and Zhang, 1995; Xie and Goh, 1997; Hoffman, 2003; and Schwertman, 2005). For a comprehensive overview of CUSUM charts for numerous distributions, see Hawkins and Olwell (1998).

¹ anwer@unizwa.edu.om; anwer_khurshid@yahoo.com

² abc_sac@rediffmail.com

In many cases, however, the entire distribution of counts is not observed. In particular, more often the zeros are not observed or sometimes a large number of zeros are contained in the data. In recent years, researchers have provided new complementary models which are obtained by modifying the existing well known models. These complementary models are generally divided into two categories; zero-truncated and zero-inflated.

Zero-truncated models are the ones when the number of individuals falling into zero class can not be determined, or the observational apparatus becomes active only when at least one event occurs. Chakraborty and Kakoty (1987) and Chakraborty and Bhattacharya (1989, 1991) have constructed CUSUM charts for zero-truncated Poisson distribution, doubly truncated GD and doubly truncated binomial distribution respectively. Recently, Chakraborty and Khurshid (2011, 2012) have constructed CUSUM charts for zero-truncated binomial distribution and doubly truncated binomial distribution respectively. Inflation occurring at any of the support point and zero-inflation indicates that a data set contains an excessive number of zeros. Chen et al. (2008) have proposed generalized zero-inflated Poisson distribution to construct attribute control chart. See also Xie et al. (2001).

Accordingly, distributions of negative binomial type often arise in practice where zero group is truncated. The main objective of this paper is to construct control charts for zero-truncated negative binomial distribution (ZTNBD) and its special case zero-truncated geometric distribution (ZTGD). Cumulative sum (CUSUM) control charts proposed by Page (1954, 1961) have been constructed for controlling the parameters of the above distributions. Control charts based on these truncated distribution are studied and Average Run Length (ARL) computed accordingly alongside developing different expressions.

2. ZERO-TRUNCATED NEGATIVE BINOMIAL DISTRIBUTION (ZTNBD)

We consider a negative binomial distribution truncated at $x = 0$. The probability mass function of the ZTNBD is

$$f(x; k, p) = \frac{\binom{x+k-1}{x} p^k q^x}{1-p^k} \quad (2.1)$$

where $x=1, 2, \dots, n$. Here $f(x; k, p)$ denotes the probability that there are x failures preceding the k -th success in the $(x+k)$ trials. The last trial must be a success, the probability of which is p and in the remaining $(x+k-1)$ trials we must have $(k-1)$ success, the probability of which is given by the binomial probability law by the expression $\frac{(x+k-1)!}{(k-1)! x!} p^{k-1} q^x$. Therefore, by compound probability theorem, $f(x; k, p)$ is given by the product of two probabilities (Biswas, 1992) i.e.,

$$f(x; k, p) = \frac{(x+k-1)!}{(k-1)! x!} p^{k-1} q^x p = \frac{(x+k-1)!}{(k-1)! x!} p^k q^x$$

for $x=0, 1, 2, \dots, n$. More formally, suppose that a box contains np non-defective items and nq defective items. Items are drawn at random with replacement, the probability that exactly $(x+k)$ trials required to produce k non-defective items is $\frac{(x+k-1)!}{(k-1)! x!} p^k q^x$. Thus k and p are the parameters of the negative binomial distribution,

where the parameters satisfy $0 < p < 1$ and $k = 1, 2, 3, \dots$

It is conventional in statistical literature to express the negative binomial distribution given (2.1) in terms of

parameters $Q = \frac{1}{p}$ and $P = \frac{1-p}{p}$, so that $Q - P = 1$, in the following form

$$f(x; k, p) = \binom{x+k-1}{x} (1-Q^{-k})^{-1} \left(\frac{P}{Q}\right)^x \left(1-\frac{P}{Q}\right)^k \quad (2.2)$$

where $x = 1, 2, \dots$. The mean and variance of the above distribution are as follows:

$$E(X) = \frac{kP}{1-Q^{-k}}, \text{ and } V(X) = \frac{kPQ}{1-Q^{-k}} \left[1 - \left(\frac{P}{Q}\right)\right] \left\{ \left(1-Q^{-k}\right)^{-1} - 1 \right\}.$$

2.1 CUSUM Control Chart for ZTNBD

Let x_1, x_2, \dots, x_n be i.i.d. random variables, each distributed with probability mass function defined in (2.2). To test the null hypothesis $H_0 : P = P_0$ and alternative hypothesis $H_1 : P = P_1 (> P_0)$ assuming k as known, we use the likelihood ratio of (2.2) as

$$\frac{f(x_1, x_2, \dots, x_n | P_1, k)}{f(x_1, x_2, \dots, x_n | P_0, k)} = \left(\frac{1-Q_0^{-k}}{1-Q_1^{-k}}\right)^{-n} (1-Q^{-k})^{-1} \left(\frac{P_1 Q_0}{P_0 Q_1}\right)^{\sum x_i} \left(\frac{Q_0(Q_1 - P_1)}{Q_1(Q_0 - P_0)}\right)^{nk}.$$

The continuation region of the sequential probability ratio test (SPRT) discriminating between the two hypotheses is given by

$$\log\left(\frac{\beta}{1-\alpha}\right) < n \log\left(\frac{1-Q_0^{-k}}{1-Q_1^{-k}}\right) + \sum_{i=1}^n x_i \log\left(\frac{P_1 Q_0}{P_0 Q_1}\right) + nk \log\left(\frac{Q_0(Q_1 - P_1)}{Q_1(Q_0 - P_0)}\right) < \log\left(\frac{1-\beta}{\alpha}\right) \quad (2.3)$$

where α and β are the probabilities of type I and type II errors respectively.

Considering the right hand side inequality of (2.3) we get

$$n \log\left(\frac{1-Q_0^{-k}}{1-Q_1^{-k}}\right) + \sum_{i=1}^n x_i \log\left(\frac{P_1 Q_0}{P_0 Q_1}\right) + nk \log\left(\frac{Q_0(Q_1 - P_1)}{Q_1(Q_0 - P_0)}\right) < \log\left(\frac{1-\beta}{\alpha}\right)$$

which gives

$$\sum_{i=1}^n x_i \log\left(\frac{P_1 Q_0}{P_0 Q_1}\right) < \log\left(\frac{1-\beta}{\alpha}\right) - n \log\left(\frac{1-Q_0^{-k}}{1-Q_1^{-k}}\right) - nk \log\left(\frac{Q_0(Q_1 - P_1)}{Q_1(Q_0 - P_0)}\right).$$

For a very small value of β (Johnson, 1961), we have

$$\sum_{i=1}^n x_i < \frac{-\log \alpha + n \left[\log\left(\frac{1-Q_0^{-k}}{1-Q_1^{-k}}\right) + k \log\left(\frac{Q_0(Q_1 - P_1)}{Q_1(Q_0 - P_0)}\right) \right]}{\log(P_1 Q_0) - \log(P_0 Q_0)}.$$

For constructing the CUSUM chart (as shown in Figure 1) we plot the sum $S_m = \sum_{i=1}^m x_i$ against the number of observation n . Suppose that O is the last plotted point, P is the vertex and the point Q is obtained by drawing a perpendicular to the line OP . The change in the value of P from P_0 to P_1 is detected if any plotted point falls below the line PQ . In this case the parameters of the mask namely the lead distance $d = OP$ and the angle of the mask $\phi = \angle OPQ$ is given by

$$d = \frac{-\log \alpha}{\left[\log \left(\frac{1-Q_1^{-k}}{1-Q_0^{-k}} \right) + k \log \left(\frac{Q_1(Q_0 - P_0)}{Q_0(Q_1 - P_1)} \right) \right]} \quad (2.4)$$

and

$$\phi = \tan^{-1} \left[\frac{\log \left(\frac{1-Q_1^{-k}}{1-Q_0^{-k}} \right) + k \log \left(\frac{Q_1(Q_0 - P_0)}{Q_0(Q_1 - P_1)} \right)}{\log(P_1 Q_0) - \log(P_0 Q_1)} \right]. \quad (2.5)$$

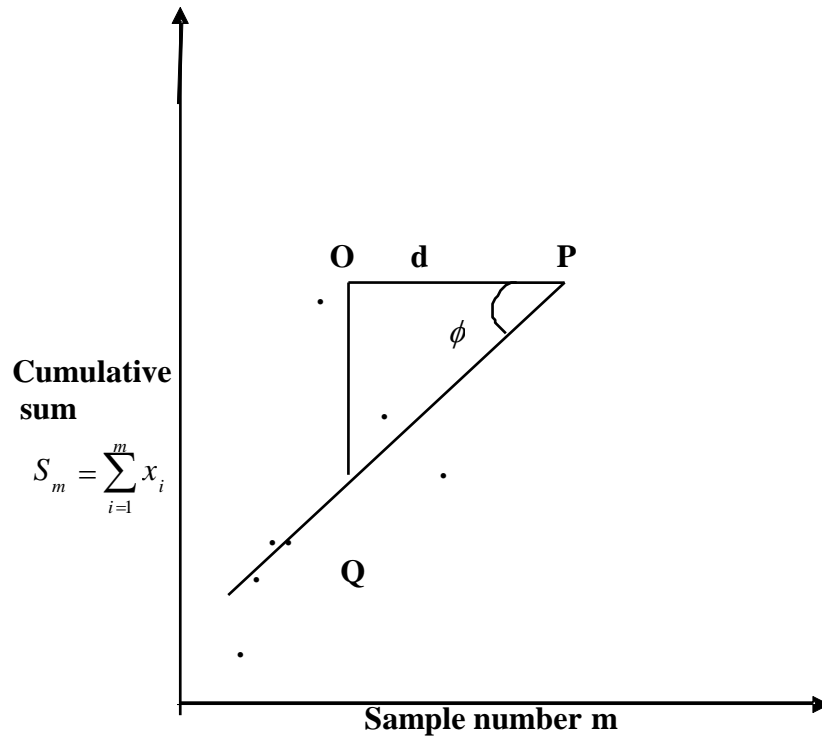


Figure 1: Cumulative sum control chart

2.2 Average Run Length (ARL)

Following Johnson (1961) (see also Johnson and Leone, 1962), the approximate formula for the average run length (ARL) for detecting a shift for the parameter P from P_0 to P_1 for known k is given by,

$$ARL = \frac{(-\log \alpha)}{E_{P_1}}$$

where $E_{P_1} = E \left[\log \frac{f(x|P_1)}{f(x|P_0)} \middle| P_1 \right]$. Thus

$$\begin{aligned} E_{P_1} &= \log \left(\frac{1 - Q_0^{-k}}{1 - Q_1^{-k}} \right) + \log \left(\frac{P_1 Q_0}{P_0 Q_1} \right) E(X) + k \log \left(\frac{Q_0(Q_1 - P_1)}{Q_1(Q_0 - P_0)} \right) \\ &= \log \left(\frac{1 - Q_0^{-k}}{1 - Q_1^{-k}} \right) + \left(\frac{k P_1}{1 - Q_1^{-k}} \right) \log \left(\frac{P_1 Q_0}{P_0 Q_1} \right) + k \log \left(\frac{Q_0(Q_1 - P_1)}{Q_1(Q_0 - P_0)} \right). \end{aligned} \quad (2.6)$$

Values of the lead distance d , angle of the v-mask and average run length (ARL) of one-sided CUSUM chart are calculated for a number of combinations of the values of P , k and α for controlling the parameters P when k is known. These values are shown in Tables 2.1 to 2.3.

Table 2.1: Values of d for controlling the parameter P

(a) When $k = 1$						
		α				
P_0	P_1	0.05	0.025	0.01	0.005	0.001
1	2	4.32	5.32	6.64	7.64	9.97
1	3	2.73	3.36	4.19	4.82	6.29
1	4	2.16	2.66	3.32	3.28	4.98
1	5	1.86	2.29	2.86	3.29	4.29

(b) When $k = 2$						
		α				
P_0	P_1	0.05	0.025	0.01	0.005	0.001
1	2	3.05	3.76	4.70	5.40	7.04
1	3	1.86	2.29	3.52	4.05	5.29
1	4	1.44	1.77	2.21	2.55	3.32
1	5	1.22	1.50	1.87	2.16	2.81

(c) When $k = 3$						
		α				
P_0	P_1	0.05	0.025	0.01	0.005	0.001
1	2	2.28	2.81	3.51	4.04	5.26
1	3	1.36	1.68	2.10	2.41	3.14
1	4	1.04	1.28	1.60	1.84	2.40
1	5	0.87	1.08	1.34	1.55	2.02

Table 2.2: Values of ϕ (in degrees) for controlling the parameter P

$k = 1$				
P_1				
P_0	2	3	4	5
1	67.47	69.72	71.28	72.40

$k = 2$				
P_1				
P_0	2	3	4	5
1	73.68	75.86	77.26	78.25

$k = 3$				
P_1				
P_0	2	3	4	5
1	77.65	79.57	80.73	81.56

Table 2.3: Values of ARL for controlling the parameter P

(a) When $k = 1$						
		α				
P_0	P_1	0.05	0.025	0.01	0.005	0.001
1	2	17.64	21.72	27.12	31.20	40.68
1	3	5.73	7.05	8.80	10.13	13.20
1	4	3.11	3.83	4.78	5.50	7.17
1	5	2.06	2.53	3.16	3.64	4.75

(b) When $k = 2$						
		α				
P_0	P_1	0.05	0.025	0.01	0.005	0.001
1	2	9.55	11.76	14.68	16.89	22.02
1	3	3.04	3.74	4.67	5.38	7.01
1	4	1.63	2.01	2.51	2.88	3.76
1	5	1.07	1.32	1.65	1.89	2.47

(c) When $k = 3$						
		α				
P_0	P_1	0.05	0.025	0.01	0.005	0.001
1	2	6.23	7.68	9.58	11.03	14.38
1	3	1.98	2.44	3.05	3.51	4.57
1	4	1.07	1.31	1.64	1.88	2.46
1	5	0.70	0.86	1.08	1.24	1.62

3. ZERO-TRUNCATED GEOMETRIC DISTRIBUTION (ZTGD)

It is known that for $k = 1$, the NBD is reduced to the geometric distribution which can be used as an alternative to the Poisson distribution for describing the number of defects or other counting data. Thus, for $k = 1$ the ZTNBD becomes zero-truncated geometric distribution (ZTGD). The probability mass function of the ZTGD for each i.i.d. random variables x_1, x_2, \dots, x_n is

$$f(x; p) = \frac{pq^x}{1-p}; \quad x = 1, 2, \dots \quad (3.1)$$

where p is the probability of success in each trial.

Suppose that we have a series of independent trials or repetitions and in each trial the probability of success p remains the same. Then the probability that there are x failures preceding the first success is given by $q^x p$ where $q = 1 - p$, and p is lying between 0 and 1.

For $f(x; k, p) = \binom{x+k-1}{x} (1-Q^{-k})^{-1} \left(\frac{P}{Q}\right)^x \left(1 - \frac{P}{Q}\right)^k$; $x = 1, 2, \dots$ the pmf of ZTGD becomes

$$f(x; p) = (1-Q^{-1})^{-1} \left(\frac{P}{Q}\right)^x \left(1 - \frac{P}{Q}\right)^k; \quad x = 1, 2, \dots \quad (3.2)$$

3.1 CUSUM Control Chart for ZTGD

The parameters of the v-mask i.e., the lead distance d and angle of the mask ϕ of the CUSUM chart (following Johnson, 1961) for ZTGD are given by

$$d = \frac{-\log \alpha}{\left[\log \left(\frac{p_0}{p_1} \right) + \log \left(\frac{1-p_1}{1-p_0} \right) \right]} \quad (3.3)$$

and

$$\phi = \tan^{-1} \left[\frac{\log \left(\frac{p_0}{p_1} \right) + \log \left(\frac{1-p_1}{1-p_0} \right)}{\log \left(\frac{1-p_1}{1-p_0} \right)} \right] \quad (3.4)$$

whereas ARL is given by

$$ARL = -\log \alpha \left[\log \left(\frac{p_1}{p_0} \right) + \log \left(\frac{1-p_0}{1-p_1} \right) + \frac{1}{p_1} \log \left(\frac{1-p_1}{1-p_0} \right) \right]^{-1} \quad (3.5)$$

For Equation (3.2), the parameters of the v-mask i.e., the lead distance d and angle of the mask ϕ are given by

$$d = \frac{-\log \alpha}{\left[\log \left(\frac{1+P_1}{1+P_0} \right) + \log \left(\frac{P_1 Q_0}{P_0 Q_1} \right) \right]} \quad (3.6)$$

and

$$\phi = \tan^{-1} \left[\frac{\log \left(\frac{1+P_1}{1+P_0} \right) + \log \left(\frac{P_1 Q_0}{P_0 Q_1} \right)}{\log \left(\frac{P_1 Q_0}{P_0 Q_1} \right)} \right], \quad (3.7)$$

and average run length (ARL) for Equation (3.2) is

$$ARL = -\log \alpha \left[\log \left(\frac{Q_1}{Q_0} \right) + \log \left(\frac{P_0 Q_1}{P_1 Q_0} \right) + Q_1 \log \left(\frac{P_1 Q_0}{P_0 Q_1} \right) \right]^{-1} \quad (3.8)$$

Tables (3.1), (3.2) and (3.3) show the values of d , ϕ and ARL for different combinations of p_0 , p_1 and α . It is to be noted that the numerical values of the Equations (3.6), (3.7) and (3.8) are the same as those shown in Tables (2.1), (2.2) and (2.3):

Table 3.1: Values of d for controlling the parameter p under ZTGD

		α				
p_0	p_1	0.05	0.025	0.01	0.005	0.001
0.2	0.3	5.56	6.84	8.54	9.82	12.82
0.2	0.4	3.05	3.76	4.70	5.40	7.04
0.2	0.5	2.16	2.66	3.32	3.82	4.98
0.2	0.6	1.67	2.06	2.57	2.96	3.86

Table 3.2: Values of ϕ (in degrees) for controlling the parameter p under ZTGD

		p_1			
p_0		0.3	0.4	0.5	0.6
1		76.09	73.65	71.27	68.85

Table 3.3: Values of ARL for controlling the parameter p under ZTGD

		α				
p_0	p_1	0.05	0.025	0.01	0.005	0.001
0.2	0.3	31.91	39.29	49.05	56.43	73.57
0.2	0.4	11.45	14.10	17.60	20.25	26.41
0.2	0.5	6.71	8.27	10.32	11.87	15.48
0.2	0.6	4.71	5.80	7.23	8.32	10.85

4. CONCLUSIONS

It is evident from the Table 2.1 that for fixed α , and fixed k , the values of d decrease as the difference $(P_1 - P_0)$ increases, whereas for the fixed difference $(P_1 - P_0)$, the values of d increases as α decreases. It is also observed from Table 2.1(a, b, c), that the values of ARL decrease as we go on increasing the values of k .

Table 2.2 shows that the angle of the mask increases as the difference between P_1 and P_0 increases and also for fixed difference, i.e., angle of the mask increases as the value of k increases. Thus from the Tables (2.1 and 2.2), we can conclude that as the lead distance d decreases, the angle of the v-mask increases where the difference is increased.

Table 2.3 shows the values of ARL (average number of observations required to detect the shift of the process parameter) for different combinations P , k and α . It is evident from the Table 2.3 (a, b and c) that for fixed α , the ARL decreases as the shift from P_0 to P_1 increases and for fixed $(P_1 - P_0)$, the ARL increases as α decreases. But for fixed change and α , the ARL decreases if k increases.

Table 3.1 shows the values of d of the Equation (3.3) (Though numerically, the calculated values are negative, but as distance can not be negative, we have considered the positive values of the distance d): Here, it is observed that for fixed α , the values of d decrease as the difference $(p_1 - p_0)$ increases, whereas for fixed difference $(p_1 - p_0)$ the values of d increase as α decreases.

From the Table 3.2 it is observed that as the difference $(p_1 - p_0)$ increases, the angle ϕ the mask decreases.

Table 3.3 shows that for fixed α , the value of ARL decreases as the difference between p_1 and p_0 increases whereas for fixed difference the value of ARL increases as α decreases.

RECEIVED NOVEMBER, 2012
REVISED FEBRUARY, 2013

REFERENCES

- [1] BISWAS, S. (1992): **Topics in Statistical Methodology**. Wiley Eastern Limited, New Delhi.
- [2] CHAKRABORTY, A. B. and bhattacharya, S. K. (1989): CUSUM control charts for doubly truncated geometric and Poisson distributions. pp. 509-512. **Proceedings of Quality for Progress and Development, Asian Congress on Quality and Reliability**. Wiley Eastern Limited, NewDelhi.
- [3] CHAKRABORTY, A. B. and BHATTACHARYA, S. K. (1991): Cumulative sum control chart for a doubly truncated binomial distribution. **Egyptian Statistical Journal**, 35, 119-124.
- [4] CHAKRABORTY, A. B. and KAKOTY, S. (1987): Control charts for ZTPD. **Journal of Indian Association for Productivity, Quality and Reliability**, 12, 17-25.
- [5] CHAKRABORTY, A. B. and KHURSHID, A. (2011): One-sided cumulative sum (CUSUM) control charts for the zero-truncated binomial distribution. **Economic Quality Control**, 26, 41-51.
- [6] CHAKRABORTY, A. B. and KHURSHID, A. (2012): Control charts for the doubly-truncated binomial distributions. **Economic Quality Control**, 27, 187-194.
- [7] 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.
- [8] HAWKINS, D. M. and OLWELL, D. H. (1998): **Cumulative Sum Charts and Charting for Quality Improvement**. Springer, New York.
- [9] HOFFMAN, D. (2003): Negative binomial control limits for count data with extra-Poisson variation. **Pharmaceutical Statistics**, 2, 127-132.
- [10] JOHNSON, N. L. (1961): A simple theoretical approach to cumulative sum chart. **Journal of the American Statistical Association**, 56, 835-840.
- [11] JOHNSON, N. L. and LEONE, F. C. (1962): Cumulative sum control charts: Mathematical principles applied to their construction and use, Part B. **Industrial Quality Control**, xix, 22-28.
- [12] JOHNSON, N. L., KEMP, A. W. and KOTZ, S. (2005): **Univariate Discrete Distributions**, Third Edition, John Wiley, New York.
- [13] 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.

- [14] KHURSHID, A., AGEEL, M. I. and LODHI, R. A. (2005): On confidence intervals for the negative binomial distribution. **Revista Investigación Operacional**, 26, 59-70.
- [15] LUCENO, A. and PUIG-PEY, J. (2000): Evaluation of the Run-Length Probability Distribution for CUSUM Charts: Assessing Chart Performance. **Technometrics**, 42, 411-416.
- [16] LUCAS, J. M. (1985): Counted Data CUSUM's. **Technometrics**, 27, 29-44.
- [17] MA, Y. and ZHANG, Y. (1995): Q control charts for negative binomial distribution. **Computers and Industrial Engineering**, 31, 813-816.
- [18] PAGE, E. S. (1954): Continuous inspection schemes. **Biometrika**, 41, 100-115.
- [19] PAGE, E. S. (1961): Cumulative sum charts. **Technometrics**, 3, 1-9.
- [20] RYAN, T. P. (2011): **Statistical Methods for Quality Improvement**, Third Edition, John Wiley, New York.
- [21] SCHWERTMAN, N. C. (2005): Designing accurate control charts based on the geometric and negative binomial distribution. **Quality and Reliability Engineering International**, 21, 743-756.
- [22] 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 and Management**, 16, 64-73.
- [23] XIE, M., HE, B. and GOH, T. N. (2001): Zero-inflated Poisson model in statistical process control. **Computational Statistics and Data Analysis**, 38, 191-201.