Nonconforming Control Charts for Zero-Inflated Poisson distribution

N. Katemee, T. Mayureesawan

Abstract—This paper developed the c-Chart based on a Zero-Inflated Poisson (ZIP) processes that approximated by a geometric distribution with parameter \( p \). The \( p \) estimated that fit for ZIP distribution used in calculated the mean, median, and variance of geometric distribution for constructed the c-Chart by three difference methods. For the c-J-Chart, developed c-Chart by used the mean and variance of the geometric distribution constructed control limits. For the cZIP-Chart, developed control limits of c-Chart from median and variance values of geometric distribution. The performance of charts considered from the Average Run Length and Average Coverage Probability. We found that for an in-control process, the cJ-Chart is superior for low level of mean at all level of proportion zero. For an out-of-control process, the cme-Chart and cme-Chart are the best for mean = 2, 3 and 4 at all level of parameter.

Keywords—average coverage probability, average run length, geometric distribution, zero-inflated poisson distribution

I. INTRODUCTION

The classical Shewhart control chart of nonconformities (c-Chart) is used to monitor the number of nonconformities per unit of product base on Poisson distribution when the sample sizes are constant. In situation, the sampling of products takes place as a set of repeated samplings, with each sampling finding either zero nonconformities or a nonzero number of nonconformities. If in some sampling processes an excess number of zeros might be observed, then distribution is called a “Zero-Inflated Poisson (ZIP)”. In this case, the estimation of the sample mean often underestimates of the Poisson distribution. When the estimated variance is greater than the mean, this is so called “Over Dispersion”, and the estimated limits in the c-Chart are improperly narrow [7]. Cohen [1] studied the maximum likelihood estimator (MLE) \( \hat{\lambda} \) of a ZIP model because he found that the estimated value for the mean by using the MLE value is closer to the actual value. This study has been used in many applications [3], [8]. Xie et al. [7] developed a c-Chart for the ZIP model that they called the cZIP-Chart. They studied the efficiency of the cZIP-Chart for monitoring upward shifts of the mean value of number of nonconformities in a process. Sim and Lim [2] illustrate a charting method called a cJ-Chart in which they used a one-sided Jeffreys prior interval [9] to detect upward shifts. They compared this cJ-Chart with a usual c-Chart and a cZIP-Chart.

They showed that the cJ-Chart was appropriate for processes when the mean was in-control. On the other hand, if the process mean was in an out-of-control situation, then the c-Chart performed better than the other charts. However, they found that the c-Chart yields poor coverage probability. Peerajit and Mayureesawan [10] extended the research ideas of [2] by supplying both the proportion of zero nonconformities and the mean shift in a production process. The results obtained for the performance of the c-Chart, cZIP-Chart and cZIP-Chart were in agreement with those of Sim and Lim.

The results of the research mentioned, the authors have either studied the performance of the charts for either the average run length (ARL) or the average coverage probability (ACP) but not both, or if they studied both they found that charts might perform well for one measure but poorly for the other measure.

The aims of the present study are to develop modified the control limits of a c-Chart for the ZIP distribution that perform satisfactorily for a range of parameters of the ZIP distribution and to compare the performance of these charts with the charts mentioned. The outline of the paper is as follows. We first develop an approximation for the distribution of the ZIP distribution as a geometric distribution with parameter \( p \), and we examine how the value of \( p \) varies as the parameter of the ZIP distribution is changed. We then use the \( p \) estimated for calculated the mean, median, and variance of geometric distribution for constructed the c-Chart by three different procedures to develop three control charts, which we call cJ-Chart, cZIP-Chart and cZIP-Chart. The performance of these developed control charts is then compared with the performance of c-Chart, cZIP-Chart and cZIP-Chart.

II. MATERIALS AND METHODS

Zero-Inflated Poisson (ZIP)

The probability mass function is given by [8]:

\[
P(Y = y) = \begin{cases} 
\frac{\omega(1-\omega)\exp -\lambda}{y!} & , \quad y = 0 \\
\frac{(1-\omega)\exp -\lambda \exp(\lambda y)}{y!} & , \quad y \geq 0, 
\end{cases}
\]  (1)

where \( Y \) = the random variables of nonconformities in a sample unit,  
\( \omega \) = the mean of nonconformities in a sample unit,  
\( \lambda \) = is a measure of the extra proportion of zero nonconformity in a sample unit, and  
\( E(Y) = (1-\omega) \mu \) and \( \mu = \left( \frac{\omega}{1-\omega} \right)^{\lambda} \).  

Note: \( \omega \) = is the Poisson distribution.

The Geometric distribution

The probability function is given by [6]:

\[
P(Y = k) = (1 - p)^{k-1} p, \quad k = 0, 1, 2, 
\]  (3)

International Scholarly and Scientific Research & Innovation 6(9) 2012 1349 scholar.waset.org/1999.7/11942
where $Y$ = the random variables of the number of failures until the first success to occur, 
$p = p_y$ = the probability of success on each trial, and 
$$E(Y) = \frac{1 - p_y}{p_y} \quad \text{and} \quad V(Y) = \frac{1 - p_y}{p_y^2}. \quad (4)$$

The Shewhart control chart of nonconformities (c-Chart)
The control limits are given by [4]:
$$\begin{align*}
UCL &= c + 3\sqrt{c} \\
CL &= c \\
LCL &= c - 3\sqrt{c}
\end{align*} \quad (5)$$
c is assumed to be the mean number of nonconformities if the mean of the probability distribution is known, otherwise $c$ is estimated as the mean of the number of nonconformities in a sample of observed product units ($\bar{c}$).

The control chart of nonconformities with ZIP model ($c_{ZIP}$-Chart)
In 1991, Cohen [1] developed a ZIP model for a Poisson probability function $g(y; \lambda, \alpha)$, $y = 0, 1, 2, \ldots$ given by:
$$P(Y = y) = \alpha^y e^{-\alpha} \left(1 - \alpha e^{-\lambda}\right)^{y-1}, \quad y = 0, 1, 2, \ldots \quad (6)$$
where $Y$ = the random variables of nonconformities in a product process, and $I_{(y,0)} = 1$ if $y = 0$ and $I_{(y,0)} = 0$ if $y \neq 0$.
The maximum likelihood estimate (MLE) of parameter $\lambda$ in the ZIP model of Cohen is given by:
$$\hat{\lambda} = \bar{y} - 1 \left(1 - e^{\frac{-1}{\lambda}}\right), \quad (7)$$
where $\bar{y}$ = the mean of the number of nonconformities in product units that have a nonzero number of nonconformities.
The $\hat{\lambda}$ are then used in the control limits for the $c_{ZIP}$-Chart [7] as follows:
$$\begin{align*}
UCL &= \hat{\lambda} + 3\sqrt{\hat{\lambda}} \\
CL &= \hat{\lambda} \\
LCL &= \hat{\lambda} - 3\sqrt{\hat{\lambda}}
\end{align*} \quad (8)$$
The control chart of nonconformities with Jeffreys Prior Interval method ($c_j$-Chart)
The one-sided Jeffreys prior interval is given by [9]:
$$CJ_j(y) = G[\alpha; y + 0.5, 1, \infty], \quad (9)$$
where $y$ = the number of nonconformities for a Poisson distribution, 
$\lambda$ = the parameter estimate $\hat{\lambda}$ for the ZIP model from (7).

If $y = 0$ the confidence interval is $[0, \infty)$, if $y \neq 0$ the confidence interval is $[G[\alpha; y + 0.5, 1, \infty]]$, where $G[\alpha; a, b]$ is the 100$\alpha$th percentile of a Gamma distribution with shape parameter $a = y + 0.5$ and scale parameter $b = 1$.

For the control chart of nonconformities with Jeffreys Prior Interval method, i.e., the $c_j$-Chart, the control limit is given by [2]:
$$UCL = \max[y \mid \lambda > G[\alpha; y + 0.5, 1]]. \quad (10)$$

Development of the $c$-Charts for Zero-Inflated Poisson (ZIP) processes
For a given ZIP distribution and $k$ is geometric distribution, we first obtain an approximate geometric distribution with parameter $p$ ($p_y$) by using the Kolmogorov-Smirnov test [5].
The charts for a ZIP process are then defined as follows:
1. $c_{ZIP}$-Chart is a modified control limits of the $c$-Chart obtained by using the $\hat{p}_y$ that fit for ZIP distribution for calculated the mean and variance of geometric distribution for constructed the control limits of a one-sided $c$-Chart. Therefore the control limit of $c_{ZIP}$-Chart is given by:
$$\begin{align*}
UCL &= E(Y) + 3\sqrt{E(Y)} \\
LCL &= 0,
\end{align*} \quad (11)$$
where
$$\hat{p}_y = \left(1 + \frac{1}{n} \sum_{i=1}^{n} k_i\right)^{-1}, \quad E(Y) = \frac{1 - \hat{p}_y}{\hat{p}_y} \quad \text{and} \quad V(Y) = \frac{1}{\hat{p}_y^2}. \quad (12)$$
2. $c_{mg}$-Chart is a modified control limits of the $c$-Chart obtained by using the $\hat{p}_y$ that fit for ZIP distribution for calculated the mean of geometric distribution for constructed the control limits of a one-sided $c$-Chart. Therefore the control limit of $c_{mg}$-Chart is given by:
$$\begin{align*}
UCL &= E(Y) + 3\sqrt{E(Y)} \\
LCL &= 0,
\end{align*} \quad (13)$$
3. $c_{me}$-Chart is a modified control limits of the $c$-Chart obtained by using the $\hat{p}_y$ that fit for ZIP distribution for calculated the median ($M$) and variance of geometric distribution for developed control limits of a one-sided $c$-Chart. Therefore the control limit of $c_{me}$-Chart is given by:
$$\begin{align*}
UCL &= M + 3\sqrt{E(Y)} \\
LCL &= 0,
\end{align*} \quad (14)$$
where $Median = \left[-\frac{1}{log_2(1 - \hat{p}_y)}\right] - 1$.

III. SIMULATION RESULTS

In this section we have shown the results of tests of the charts by a simulation study. For the simulations, we assume the following ranges of parameter values. The means for the in-control process are: $(\mu_0) = 1.0(1.0)4.0$. The means for the out-of-control process are: $(\mu_1 = \mu_0 + \rho)$ where the mean shifts are: $(\rho) = 0.00, 0.40, 0.80$ and $1.20$. The proportions of zero nonconformity are: $(\omega) = 0.30(0.10)0.90$. Finally, the value for the over-dispersion $(\varphi) = 1$.

The evaluation of the performance of the control charts was conducted as follows:
1. The R program was used to simulate the number of nonconforming items for a ZIP distribution with values for the parameters $(n, \mu_0, \varphi, \omega)$ chosen from the set of values given above.
2. The value of the parameter \( p_z \) which gives a best fit between the ZIP distribution from step 1 and geometric distribution.

3. The Kolmogorov-Smirnov test was used to test the hypothesis that a geometric distribution with the \( p_z \) value from step 2 could give a reasonable fit to the distribution of data obtained in step 1. Based on simulations with 20,000 replications, the results of the test showed that the hypothesis was satisfied for at least 95% of the replications. For the \( p_z \) fit values with a ZIP distribution, we used the number of failures until the first success to occur of a geometric distribution for calculated the \( \hat{p}_z \) from (12).

4. Based on 100,000 replications, the averaged control limits were calculated for the \( c-Chart \), \( cZIP-Chart \) and \( cJ-Chart \). For the \( c ZIP-Chart \), \( cmg-Chart \) and \( cmc-Chart \) the values for the average \( \hat{p}_z \) in step 3 were then used for calculating the mean, median, and variance of geometric distribution for construct the control limits.

5. Based on a new set of 100,000 replications, the control limits calculated in step 4 were then used to compute the \( ARL \) and the \( ACP \) for each chart.

6. Steps 1 to 5 were then repeated for a new set of values for parameters \( (n, \mu_0, \varphi, \omega) \).

IV. RESULTS

In this section a summary is given of some of the results that were obtained from the simulations.

Table I shows the values of \( \hat{p}_z \) for the geometric distribution that gives the best fit between the geometric and the ZIP distribution for a range of \( \omega \) and \( \mu \) values. It can be seen that as the values of \( \mu = 1.0 \), the values of \( \hat{p}_z \) for all of \( \omega \) are a constant value (0.53) and when the values of \( \mu = 2.0 - 4.0 \), as the values of \( \omega \) are increased, the values of \( \hat{p}_z \) vary depend on the \( \mu \) values.

The results for the in-control case (\( \rho = 0.00 \)) are shown in Table II. Table II shows a comparison of \( ARL_0 \) and \( ACP \) values for the \( c-Chart \) \((c)\), \( cZIP-Chart \)(\(cZIP\)), \( cJ-Chart \)(\(cJ\)), \( cmc-Chart \)(\(cmc\)) and \( cmg-Chart \)(\(cmg\)). A comparison of \( ARL_0 \) values for the charts is given in Fig. 1. It can be seen that when \( \mu_0 = 1.0 \) for all levels of \( \omega \) the \( cZIP-Chart \) and \( cJ-Chart \) return similar the highest \( ARL_0 \) values. Therefore, they are accepted as the preferred control chart because it detects shifts slowly. When \( \mu_0 = 2.0 - 4.0 \) and for \( \omega = 0.3 \) and 0.4, the \( cZIP-Chart \) returns the highest \( ARL_0 \) values. Therefore \( cJ-Chart \) is appropriate control chart. However, when \( \omega = 0.5 - 0.9 \), the \( cZIP-Chart \) is appropriate control chart. Fig. 2 shows the absolute values of the differences between the \( ACP \) values and the confidence level of 0.9973, which we call the \( ACP-DIFF \) value, for the preferred charts for the \( ARL \) values i.e., for the \( cJ-Chart \) and \( cZIP-Chart \). It can be seen that when \( \omega = 0.3 \) and 0.4, these two charts have similar low \( ACP-DIFF \) values for all values of \( \mu_0 \).

That is, these control charts all give \( ACP \) values close to the target level of 0.9973. However, for higher \( \omega \) (0.5–0.9), only the \( cZIP-Chart \) and \( cJ-Chart \) give \( ACP \) values close to the required confidence level. When both \( ARL_0 \) and \( ACP \) values are considered, the \( cZIP-Chart \) and \( cJ-Chart \) will be the preferred control chart when \( \omega = 0.3 \) and 0.4 for levels of \( \mu_0 = 1.0 \). However, only the \( cJ-Chart \) will be the preferred control chart for \( \mu_0 = 2.0 - 4.0 \). When all of \( \mu_0 \) and for \( \omega = 0.5 - 0.9 \), only the \( cJ-Chart \) will be the preferred control chart.

The process is in an out-of-control state (\( \rho > 0.00 \)).

Results for this case are shown in Fig. 3 and 4. Fig. 3 gives a comparison of \( ARL \) values for a range of values of \( \mu_0 = \mu_0 + \rho \), \( \omega \), and \( \rho \). It can be seen that when \( \mu_0 = 1.0 \) for \( \omega = 0.3 - 0.6 \) and all of \( \rho \), the \( c-Chart \), \( cmc-Chart \) and \( cmg-Chart \), return similar low values of \( ARL \). That is, these three charts are able to detect shifts faster than the other charts. However, only the \( cJ-Chart \) return low values for \( \omega = 0.7 - 0.9 \). When \( \mu_0 = 4.0 \) (as \( \mu_0 = 2.0 \) and 3.0 return similar result) for all of \( \omega \) and \( \rho \), the \( c-Chart \), \( cmc-Chart \) and \( cmg-Chart \) return similar low values of \( ARL \). It can be seen from Fig. 3, that all control charts detect a shift slowly for values of \( \omega \) of (0.8, 0.9). Fig. 4 gives a comparison of the \( ACP-DIFF \) values for the preferred charts for the \( ARL \) values, that is, for the \( cZIP-Chart \) and \( cmc-Chart \). It can be seen that when all of \( \omega \), the \( cmc-Chart \) and \( cmg-Chart \) return the lowest \( ACP-DIFF \) values for all values of \( \mu_0 \) and \( \rho \), as \( \mu_0 = 1.0 \) and \( \rho = 0.3-0.6 \) and all of \( \rho \).

However, no control charts are to be preferred for \( \omega = 0.7 - 0.9 \). When levels of \( \mu_0 = 2.0 - 4.0 \) for all \( \omega \) and \( \rho \), the \( cmc-Chart \) and \( cmg-Chart \) will be the preferred control chart.

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>N ( \mu_0 ) VALUES FOR THE GEOMETRIC THAT GIVE THE BEST FIT TO THE DISTRIBUTION OF THE ZIP MODEL FOR A RANGE OF ( \mu ) AND ( \omega ) VALUES</td>
</tr>
<tr>
<td>( \mu )</td>
</tr>
<tr>
<td>1.0</td>
</tr>
<tr>
<td>0.30</td>
</tr>
<tr>
<td>0.40</td>
</tr>
<tr>
<td>0.50</td>
</tr>
<tr>
<td>0.60</td>
</tr>
<tr>
<td>0.70</td>
</tr>
<tr>
<td>0.80</td>
</tr>
<tr>
<td>0.90</td>
</tr>
</tbody>
</table>
### Table II
Comparison of ARL₀ and ACP values of the C-Chart, CZIP-Chart, C_J-Chart, C_g-Chart, CMG-Chart and CME-Chart for a range of \( \mu_0 \) and \( \omega \) values

<table>
<thead>
<tr>
<th>( \mu_0 )</th>
<th>ARL₀</th>
<th>ACP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( c )</td>
<td>( c_{ZIP} )</td>
</tr>
<tr>
<td>1.0</td>
<td>0.3</td>
<td>74.5</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>19.8</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>23.8</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>30.1</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>11.7</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>17.9</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>14.7</td>
</tr>
<tr>
<td>2.0</td>
<td>0.3</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>13.1</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>16.3</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>14.4</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>15.8</td>
</tr>
<tr>
<td>3.0</td>
<td>0.3</td>
<td>41.7</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>18.7</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>12.6</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>11.4</td>
</tr>
</tbody>
</table>
Fig. 2 Comparison of $ACP-DIFF$ values of the $cZIP$-Chart, $cJ$-Chart and $cg$-Chart for a range of $\mu_0$ and $\omega$ values.

Fig. 3 Comparison of $ARL_1$ values of the $c$-Chart, $cZIP$-Chart, $cJ$-Chart, $cg$-Chart, $cmg$-Chart and $cme$-Chart for a range of $\mu_0 = 1.0$, $\rho = 0.4$, $\mu_0 = 1.0$, $\rho = 1.2$, $\mu_0 = 4.0$, $\rho = 0.4$ and $\mu_0 = 4.0$, $\rho = 1.2$ values.
V. CONCLUSION

In this paper, three control charts have been proposed for a process with number of non-conformities from a ZIP distribution. In developing the charts, the number of non-conformities is modeled as a geometric distribution with parameter $p_g$, where $p_g$ estimated gives the best fit between the geometric and ZIP distributions used in calculated the mean, median, and variance of geometric distribution for constructed the c-Chart by three difference methods. The three charts are called the $c_g$-Chart, $cmg$-Chart and $cme$-Chart. In the $c_g$-Chart, we constructed the control limits with the mean and variance of geometric distribution. In the $cmg$-Chart, we used the median and variance of geometric for replaced in control limits. In the $cme$-Chart, we used the median and variance of geometric for replaced in control limits.

Additionally, the simulations have been carried out to compare the performances of the three control charts with the performances of three other charts: c-Chart, cZIP-Chart and cJ-Chart. The average run length (ARL) and average coverage probability (ACP) have been compared. The results of the comparisons are summarized in table III which gives a list of preferred control charts for both in-control and out-of control states for a range of values of ZIP parameters.

<table>
<thead>
<tr>
<th>The mean shift of process</th>
<th>Mean of process ($\mu_g / \mu$)</th>
<th>Proportion of zero ($\omega$)</th>
<th>Preferred control charts</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.3 - 0.4</td>
<td>$c_g$-Chart and $c_f$-Chart</td>
<td>$c_g$-Chart and $c_f$-Chart</td>
</tr>
<tr>
<td></td>
<td>0.5 - 0.9</td>
<td>$c_g$-Chart and $c_f$-Chart</td>
<td>$c_f$-Chart</td>
</tr>
<tr>
<td>2.0 - 4.0</td>
<td>0.3 - 0.4</td>
<td>$c_g$-Chart and $c_f$-Chart</td>
<td>$c_f$-Chart</td>
</tr>
<tr>
<td></td>
<td>0.5 - 0.9</td>
<td>$c_g$-Chart and $c_f$-Chart</td>
<td>$c_f$-Chart</td>
</tr>
<tr>
<td>Out-of-control</td>
<td>(all level of $\mu$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.3 - 0.6</td>
<td>c-Chart, cmg-Chart and cme-Chart</td>
<td>cmg-Chart and cme-Chart</td>
</tr>
<tr>
<td></td>
<td>0.7 - 0.9</td>
<td>c-Chart</td>
<td>cmg-Chart and cme-Chart</td>
</tr>
<tr>
<td></td>
<td>2.0 - 4.0</td>
<td>c-Chart, cmg-Chart and cme-Chart</td>
<td>cmg-Chart and cme-Chart</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENT

The authors would like to thank for Rajamangala University of Technology Lanna for the financial support during this research.
REFERENCES


