MultiCraft

International Journal of Engineering, Science and Technology
Vol. 13, No. 3, 2021, pp. 37-48

**INTERNATIONAL JOURNAL OF ENGINEERING, SCIENCE AND TECHNOLOGY**

www.ijest1-ng.com
www.ajol.info/index.php/ijest


# Measuring capabilities of zero-inflated processes

**Surajit Pal [1], Susanta Kumar Gauri [2]***

*[1] SQC & OR Unit, Indian Statistical Institute, 110, N. Manickam Road, Chennai-600029, INDIA*
*[2] SQC & OR Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata-700108, INDIA*
**Corresponding Author: e-mail: susantagauri@hotmail.com, Tel. No.: 091-033-2575-3350; Fax No.: 091-033-2577-6042*
*ORCID iDs: https://orcid.org/ 0000-0001-6066-5782 (Pal), https://orcid.org/ 0000-0003-4654-7696 (Gauri)*

**Abstract**

The high-quality processes usually have more count of zeros than are expected under chance variation of its underlying Poisson or other count distribution. Therefore, these processes are usually referred to as zero-inflated processes. The zero-inflated processes are commonly modelled by zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB) distribution. In a manufacturing set up, the evaluation of process capability index of a zero-inflated process can be useful in many ways, e.g. i) predicting how well the process will hold the specifications, ii) selecting between competing vendors, and iii) assisting product developers/designers in modifying the process, etc. However, researchers have given very little attentions on this aspect of zero-inflated processes. Only one such attempt is reported in literature. But, it does not always represent the true capabilities of zero-inflated processes, and sometimes it may give very misleading impression about the capability of the concerned process. In this article, the concept of Borges and Ho (2001) is applied to zero-inflated processes and a new approach for computation of process capability index of zero-inflated processes is developed. The proposed method reveals the true capabilities of zero-inflated processes consistently. Application of the proposed approach and its effectiveness are illustrated using two datasets published by past researchers.

*Keywords:* Count data, process capability index, zero-inflated process, zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB).

DOI: http://dx.doi.org/10.4314/ijest.v13i3.4

*Cite this article as*:
Pal S., Gauri S.K. 2021. Measuring capabilities of zero-inflated processes. *International Journal of Engineering, Science and Technology*, Vol. 13, No. 3, pp. 37-48. doi: 10.4314/ijest.v13i3.4

Received: July 22, 2021; Accepted: August 17, 2021; Final acceptance in revised form: August 20, 2021

## 1. Introduction

Statistical quality control (SQC) plays an important role in many manufacturing and service industries, and it includes primarily the areas of acceptance sampling, statistical process control (SPC), and capability evaluation. Acceptance sampling uses statistical sampling to determine whether to accept or reject a production lot of material. Dodge (1943) presented the pioneering work on acceptance sampling. The fundamentals of SPC and control charting were proposed by Walter Shewhart in the 1920s and 1930s. One of the main purposes of control charts is to identify the variation due to assignable causes so that appropriate corrective measures can be taken in the manufacturing process in order to keep the process as stable under the influence of chance causes of variation alone. The process capability is a utility measure which indicates the total amount of variation in a stable process. The concept of $C_p$, the first process capability index, is introduced by Kane (1986). Process capability index provides single number assessment of the ability of the process to meet specification limits for the quality characteristics of interest. Subsequently, many other indices ($C_{pk}$, $C_{pm}$, $C_{pmk}$ etc.) are developed to overcome the limitations of $C_p$. Over the years, there were many important advances in all the areas of SQC and these advancements are well documented in different books on SQC, e.g. Montgomery (2012) and Polhemus (2018).

The quality revolution caused by an increasingly competitive global market since 1990s coupled with the rapid advancement of technologies and automation in today's world has led to tremendous improvement in the quality of manufactured products, where process performance is measured in terms of number of defectives per million units instead of percentage of defectives. In such processes, majority of the produced items are defect free and only a bare minimum number of items are observed as defective items containing single/multiple types of defects. One assumption is that these processes are so good that, in general, the produced items are defect-free but the process is subject to random shocks (Chang and Gan, 1999; Xie and Goh, 1993). The random shocks cause occurrences of some defects (or defectives) and the number of defects (or defectives) follows a Poisson (or binomial) distribution. Such high-quality processes usually have more count of zeros than are expected under chance variation of its underlying Poisson or other count distribution (Gupta et al., 1996; Sim and Lim, 2008). Therefore, these processes are usually referred to as zero-inflated processes (Lambert, 1992; Sim and Lim, 2008).

The traditional techniques of SQC, e.g. $p$-chart, $c$-chart and $u$-chart for process monitoring, published sampling plans for lot sentencing, available methodologies for process capability evaluation become no more applicable to the zero-inflated processes. This is because Poisson distribution or binomial distribution fails miserably to model sample count data obtained from zero-inflated processes where most of the items are defect free. Since 1990s researchers have taken substantial interests in zero inflated processes and attempted to develop appropriate techniques that can be applied effectively to zero inflated processes.

The aspect of process control and monitoring of zero inflated processes has drawn maximum attention of the researchers. Xie and Goh (1993), Xie et al. (2001) and Sim and Lim (2008) advocated to fit a zero-inflated Poisson (ZIP) model to account for the excess number of zeros and then to determine the control limits of the control charts using the parameter value estimated from the fitted model. Chang and Gan (1999) developed control charts based on zero-inflated Geometric (ZIG) distribution. Sim and Lim (2008) constructed control chats using the parameter values estimated from the fitted zero-inflated binomial (ZIB) model to the observed count data. Alevizakos and Koukouvinos (2020) proposed double exponentially weighted moving average (DEWMA) control chart for monitoring of zero-inflated binomial processes. Rakitzis et al (2016) have proposed CUSUM control charts for monitoring ZIB processes. Mahmood and Xie (2019) have reviewed in details the past and current trends for the models and monitoring of zero-inflated processes. Considerable researches are also carried out, in the recent past, on determination of appropriate acceptance sampling plans under different zero inflated count distributions (Loganathan and Shalini, 2014; Rao and Aslam, 2017).

However, the problem of assessment of capabilities of zero inflated processes has taken very little attention of the researchers. Traditionally, capabilities of processes are assessed in terms of different indices, e.g. $C_p$, $C_{pk}$, $C_{pm}$, and $C_{pmk}$ (Chen at al., 2017; Kane, 1986; Kotz and Johnson, 2002). Historically, these indices are developed for a product characteristic which can be described as a continuous variable that follows normal distribution. The generalization of these indices for continuous non-normal quality characteristics are suggested by Clements (1989), Pearn and Chen (1995) and many others. But, in reality, there exists many quality characteristics which are neither continuous variable, nor do they follow normal distribution. These data (e.g. defect, error etc.) are typically obtained by counting and known as attribute (count) data, which usually follow Poisson or binomial distribution. Therefore, standard formulas cannot be used for computation of capability indices of a process involving such characteristics. To alleviate the problem, some generalized indices, e.g. $C_f$ index (Yeh and Bhattacharya, 1998), $C$-index (Borges and Ho, 2001), $C_{pc}$ index Perakis and Xekalaki, 2005) and $C_{py}$ index (Maiti et al., 2010) are proposed in literature. These indices are applicable to any process regardless of whether the quality characteristic is discrete or continuous and its underlying probability distribution. The attribute characteristics are usually smaller-the-better (STB) type and having only upper specification limit. Thus, the appropriate generalized indices for these characteristics are $C_{fu}$, $C_u$, $C_{pcu}$ and $C_{pyu}$. Pal and Gauri (2020[a], 2020[b]) have compared the relative accuracies of these one-sided generalized indices for binomial as well as Poisson processes.

In a manufacturing set up, the evaluation of process capability index of a zero-inflated process can be useful in many ways, e.g. i) predicting how well the process will hold the specifications, ii) selecting between competing vendors, and iii) assisting product developers/designers in modifying the process, etc. So, process capability analysis of zero-inflated process has an important role in the context of quality control. However, researchers have given very little attentions on this aspect of zero-inflated processes. To the best of our knowledge, only Patil and Shirke (2012) have attempted to measure capability of a zero-inflated process. They have incorporated the inflation of zero ($\pi$) parameter into $C_{pcu}$ index, proposed by Perakis and Xekalaki (2005), and denoted the new index as $C_{pcu}^{Z}$. However, it fails to represent the true capabilities of zero-inflated processes consistently. Particularly, for small value of $\pi$ (say $\leq 0.2$), the value of $C_{pcu}^{Z}$ index becomes unusually high, which gives a wrong impression about the capability of the concerned process.

In this article, the concept of Borges and Ho (2001) is applied to zero-inflated processes and a new approach for computation of process capability index for zero-inflated processes is developed, which reveals the true capabilities of zero-inflated processes consistently. In section 2, we discuss about the two most commonly used models (ZIP and ZINB distributions) for modelling of zero-inflated count data. In section 3, the procedures for estimating parameters of these distributions and selection of the most appropriate distribution for describing the concerned zero-inflated count data are discussed. The proposed approach for assessment of capability of a zero-inflated process is discussed in section 4. Analysis of two datasets published by past researchers and the related results are presented in section 5 for illustrations of the proposed approach and its effectiveness. Section 6 concludes the paper.

## 2. Modelling Zero-inflated Count Data

A sample of size $n$ collected from a zero-inflated process always contains more count of zeros than are expected under chance variation of its underlying standard distribution. This extra zeros cause overdispersion (i.e. variance be larger than the mean) and thus, modifications in the underlying standard distributions are needed to avoid the incorrect estimation of the model parameters and standard errors. Zero inflated Poisson (ZIP) model is usually used in modelling zero-inflated count data where the overdispersion is solely caused by the extra zeros. Test procedures for checking zero-inflation are available in Zhao et al. (2009) and Kumar and Ramachandran (2020). Yang et al. (2011) have proposed a method for outlier identification and robust parameter estimation in a zero-inflated Poisson process. For count data where the overdispersion is caused by excess zeros and also by unobserved heterogeneity, the most commonly recommended model is zero inflated negative binomial (ZINB). This is because it employs additional parameter that models additional variability (Chaney et al., 2013; Martin and Hall, 2017). Some other models that are used for such overdispersed count data are zero inflated double Poisson (ZIDP) model (Phang and Loh, 2013) and zero inflated generalized Poisson (ZIGP) (Wagh and Kamalja, 2018). Workie and Azene (2021) have proposed Bayesian zero-inflated regression model. Fávero et al. have proposed zero-inflated generalized linear mixed models. All these models are developed by assuming that the outcome variable contains a mixture of a point mass at zero and a count distribution. However, the most popular zero inflated models used by many researchers are ZIP and ZINB models. So, here only these two models are considered for modelling zero-inflated processes.

### *2.1 Zero-inflated Poisson (ZIP) model*

Let us assume that only one type of random shock occurs in the zero-inflated process and the probability of occurrence of that random shock is $\Omega$ (where, $1 < \Omega < 1$). If Y is an independent random variable having a zero-inflated Poisson distribution, the zeros are assumed to occur in two ways corresponding to two distinct underlying states. The first state (random shock) occurs with probability $\Omega$ and when it occurs, the counts of defects, i.e. nonconformities in samples follow a Poisson distribution with parameter $\lambda$ (where , $\lambda > 0$), and the other state occurs with probability 1- $\Omega$. The zeros from the Poisson distribution are called sampling zeros and zeros from the second state are called structural zeros. The probability mass function (pmf) for ZIP model is given by

$$f(y;\Omega,\lambda) = \begin{cases} (1-\Omega) + \Omega e^{-\lambda} & \text{for } y = 0 \\ \Omega \frac{\lambda^y e^{-\lambda}}{y!} & \text{for } y = 1,2,3,.. \end{cases} \quad (1)$$

Here, mean and variance of the underlying Poisson distribution is $\lambda$, and the mean and variance of the ZIP distribution are $E(Y) = \Omega\lambda$ and $Var\ (Y) = \Omega\lambda[1 + (1 - \Omega)\lambda]$ respectively.

### *2.2. Zero-inflated negative binomial (ZINB) model*

For count data where the overdispersion is caused by excess zeros as well as unobserved heterogeneity, the most commonly recommended model is ZINB (Chaney et al., 2013). This is because it employs additional parameter that models additional variability. Suppose, the probability of occurrence of a random shock is $\Omega$ and when a random shock occurs, the counts of nonconformities follow a negative binomial (NB) distribution with parameters $k$ and $\lambda$, where $k$ ($> 0$) is the dispersion parameter and $\lambda$ is the mean of NB distribution. Then the probability mass function (pmf) for ZINB model is given by

$$f(y;\Omega,k,\lambda) = \begin{cases} (1-\Omega) + \Omega\left(\frac{k}{k+\lambda}\right)^k & \text{for } y = 0 \\ \Omega \times \frac{\Gamma(y+k)}{y!\Gamma(k)}\left(\frac{\lambda}{k+\lambda}\right)^y\left(\frac{k}{k+\lambda}\right)^k & \text{for } y = 1,2,3,.. \end{cases} \quad (2)$$

Here, the mean and variance of the underlying negative binomial distribution are $\lambda$ and $\lambda\left(1+\frac{\lambda}{k}\right)$ respectively. On the other hand, the mean and variance of the ZINB variable are $E(Y) = \Omega\lambda$ and $Var\ (Y) = \Omega\left[\lambda + \left(1 - \Omega + \frac{1}{k}\right)\lambda^2\right]$ respectively.

## 3. Estimating Parameters of ZIP and ZINB Distributions

Let us assume that $n$ units of products are randomly collected from a manufacturing process and numbers of nonconformities present in each of the sample units are observed. Suppose, number of units each having $'i'$ number of nonconformities, i.e. frequency of $'i'$ ($i$ = 0,1,2,3,...,$m$) number of defects (or nonconformities) in the collected sample units is denoted as $n_i$. Therefore, $\sum_{i=0}^{m} n_i = n$. The total number of nonconformities in $n$ sample units can be computed as D = $\sum_{i=0}^{m} i \times n_i$.

### *3.1 Estimation of parameters of ZIP distribution*

In the ZIP model, Ω is the probability of occurrence of a random shock and λ is the average number of nonconformities in a sample, when the shock occurs. The parameters Ω and λ can be easily estimated from the observed dataset by the method of maximum likelihood (Xie and Goh, 1993). The log-likelihood function of Ω and λ for the observed dataset can be written as

$$\ln L(\Omega, \lambda) = n_0 \, ln\left[(1-\Omega) + \Omega e^{-\lambda}\right] + \sum_{i=1}^{m} n_i ln\,(\Omega \frac{e^{-\lambda}\lambda^i}{i!}) \tag{3}$$

The partial derivatives of the log-likelihood function with respect to Ω and λ result in the following two likelihood equations:

$$\frac{n_0(-1+e^{-\lambda})}{(1-\Omega)+\Omega e^{-\lambda}} + \frac{n-n_0}{\Omega} = 0 \tag{4}$$

$$\frac{-n_0 \Omega e^{-\lambda}}{(1-\Omega)+\Omega e^{-\lambda}} + \frac{D}{\lambda} - (n - n_0) = 0 \tag{5}$$

The maximum likelihood estimates (MLEs) of Ω and λ can be determined by solving the two likelihood equations, which are complicated. Therefore, these likelihood equations need to be solved numerically. It is observed that the optimal values of Ω and λ can easily be determined by performing enumerative search using the ‘Solver’ tool of Microsoft Excel. Using the ‘Solver’ tool, the effects of all the possible values of Ω and λ (subject to the constraints due to the two likelihood equations) on the log-likelihood function can be examined and then, the values of Ω and λ that maximizes the log-likelihood function can be determined. These values of Ω and λ will be the MLEs of Ω and λ respectively.

Once the MLEs of Ω and λ are obtained, the expected proportions of units having $'i'$ ($i$ = 0,1,2,3,...,$m$) number of nonconformities and expected frequency of $'i'$ ($i$ = 0,1,2,3,...,$m$) number of nonconformities in a sample of size $n$ units can easily be determined. It is important to carry out the Chi-square goodness-of-fit test for checking the adequacy of the fitted model. If it passes the goodness-of-fit test, it can be assumed that the fitted ZIP model is appropriate for representing the distribution of nonconformities in the produced items.

### *3.2 Estimation of parameters of ZINB distribution*

The parameters (Ω, λ and $k$) of the ZINB distribution can be estimated from the observed dataset by applying maximum likelihood method. The log-likelihood function of Ω, λ and $k$ for the observed dataset can be written as under:

$$\ln L(\Omega, k, \lambda) = n_0 \times ln\left\{(1-\Omega) + \Omega\left(\frac{k}{k+\lambda}\right)^k\right\} +$$
$$\sum_{\mathrm{i}=1}^{m} n_i [ln(\Omega) + ln\{\Gamma(y_i + k)\} - ln(y!) - ln\{\Gamma(k)\} + + k\, ln(k) + y_i\, ln(\lambda) - (y_i + k) ln\,(\lambda + k)] \tag{6}$$

An iterative optimization procedure is required to determine the values of Ω, λ and $k$ that maximize the log-likelihood function. These values are the MLEs of the three unknown parameters of the ZINB distribution. It is observed that the optimal values of Ω, λ and $k$ can be determined by performing enumerative search using the ‘‘Solver’’ tool of Microsoft Excel. Taking into account the applicable constraints, the ‘‘Solver’’ tool can examine the effects of all the possible values of Ω, λ and $k$ on the log-likelihood function and thus, it can easily find out the values of Ω, λ and $k$ that maximize the log-likelihood function. In this case, the appropriate constraint is that the mean of the fitted ZINB distribution is equal to the sample mean.

Once the MLEs of Ω, λ and $k$ are obtained, the expected proportions of units having $'i'$ ($i$ = 0,1,2,3,...,$m$) number of nonconformities and expected frequency of $'i'$ ($i$ = 0,1,2,3,...,$m$) number of nonconformities in a sample of size $n$ units can easily be determined. Then, Chi-square goodness-of-fit test should be carried out for checking the adequacy of the fitted model.

### *3.3 Selection of the most appropriate model*

To a given dataset, both ZIP and ZINB distributions can be fitted, and both the fitted distributions may pass Chi-square goodness-of-fit test. However, one should use the most appropriate models for the purpose of statistical process control. For identification of the most appropriate model for a zero-inflated process generally the following two information criterion are used: i) Akaike Information Criterion (AIC) and ii) Bayesian Information Criterion (BIC). The AIC and BIC are formally defined as

$$AIC = 2K - 2 \times \ln(L) \tag{7}$$
$$BIC = K\ln(n) - 2 \times \ln(L) \tag{8}$$

where $K$ is the number of estimated parameters in the model, $\ln(L)$ is the log-likelihood function for the model, and $n$ is the sample size. A smaller value of AIC (or BIC) implies that the existing discrepancy between the fitted model and the data is less, and thus the fitted model that results in the minimum value of AIC (or BIC) can be considered as the most appropriate one.

It is important to note that AIC (or BIC) value reveals the relative goodness of two or more fitted models. But it does not say anything about the adequacy of the fitted model. Therefore, the primary criterion for considering a fitted model as the candidate model for the comparison is that it must pass the Chi-square goodness-of-fit.

## 4. Proposed Approach for Assessment of Capabilities of Zero-inflated Processes

The standard formulas for process capability indices are developed for normal processes, which are symmetric about the mean. However, count data (known as attribute data) are discrete and its distributions (usually Poisson or binomial) are not symmetric about the mean. Similarly, ZIP and ZINB distributions also are not symmetric about the mean. Therefore, standard formulas cannot be used for computation of capability indices of these processes that involve attribute quality characteristics. To alleviate the problem of evaluation of capability index of Poisson or binomial process, some generalized indices, e.g. $C_f$ index (Yeh and Bhattacharya, 1998; $C$-index (Borges and Ho, 2001), $C_{pc}$ index (Perakis and Xekalaki, 2005) and $C_{py}$ index (Maiti et al., 2010) are proposed in literature. The attribute characteristics are usually smaller-the-better (STB) type having only upper specification limit (USL), and thus, the appropriate generalized indices for one-sided specification of these characteristics are $C_{fu}$, $C_u$, $C_{pcu}$ and $C_{pyu}$. Pal and Gauri (2020[a], 2020[b]) have compared the relative accuracies of these one-sided generalized indices for binomial as well as Poisson processes. They have found that $C_u$ index (Borges and Ho, 2001) gives the most accurate estimate of the process capability for the binomial as well as Poisson processes.

The $C_u$ index (Borges and Ho, 2001) has one-to-one correspondence (mapping) between the proportion of nonconformance and Z-value of the standard normal distribution. In this method, the expected proportion of nonconformance of a characteristic above USL is mapped to the Z-score in the right side of standard normal distribution, and 1/3rd of this Z-score is considered as the measure of the process capability with respect to USL and it is denoted as $C_u$. It is proposed to use the same procedure for evaluation of capabilities of zero-inflated processes. The $C_u$ index can be evaluated from a zero-inflated process using the following procedures:

1) Collect a sample of $n$ units from the concerned zero-inflated process and observe the numbers of nonconformities present in each of the sample items. Let the random variable Y represents number of nonconformities presents in an item.
2) Fit appropriate zero-inflated distribution to the observed count data.
   The procedure for fitting of the two most commonly used zero-inflated models and subsequent selection of the most appropriate zero-inflated model is described in section 3.
3) Estimate the expected proportion of nonconforming items with respect to USL ($PNU_{USL}$) in the concerned zero inflated process. Let $c_{USL}$ be the USL specified by the manufacturer on the number of nonconformities in a unit. A unit will be considered nonconforming if the number of nonconformities in it is more than $c_{USL}$. Then $PNU_{USL}$ can be estimated as follows:
   - For ZIP process, the $PNU_{USL}$ can be estimated as
   $$\widehat{PNU}_{USL} = P(Y > c_{USL}) = 1 - P(Y \leq c_{USL})$$
   $$= 1 - \left[\left(1-\widehat{\Omega}\right) + \widehat{\Omega}e^{-\hat{\lambda}} + \sum_{y=1}^{c_{USL}} \widehat{\Omega}\frac{\hat{\lambda}^y e^{-\hat{\lambda}}}{y!}\right]$$
   $$= \left[\widehat{\Omega}\left(1-e^{-\hat{\lambda}}\right) - \sum_{y=1}^{c_{USL}} \widehat{\Omega}\frac{\hat{\lambda}^y e^{-\hat{\lambda}}}{y!}\right] \tag{9}$$
   - For ZINB process, the $PNU$ can be estimated as
   $$\widehat{PNU}_{USL} = P(Y > c_{USL}) = 1 - P(Y \leq c_{USL})$$
   $$= 1 - \left[\left(1-\widehat{\Omega}\right) + \widehat{\Omega}\left(\frac{\hat{k}}{\hat{k}+\hat{\lambda}}\right)^{\hat{k}} + \sum_{y=1}^{c_{USL}} \widehat{\Omega} \times \frac{\Gamma(y+\hat{k})}{y!\Gamma(\hat{k})}\left(\frac{\hat{\lambda}}{\hat{k}+\hat{\lambda}}\right)^y \left(\frac{\hat{k}}{\hat{k}+\hat{\lambda}}\right)^{\hat{k}}\right] \tag{10}$$
4) Determine the Z-value in the right side of the standard normal distribution that results in probability area equal to $\widehat{PNU}_{USL}$value. In other words, map the computed $\widehat{PNU}_{USL}$value to the Z-score in the right side of standard normal distribution. Let $Z_U$ is the value of Z that results in probability area $\widehat{PNU}_{USL}$ above it. The $Z_U$ value can be obtained by using inverse cumulative probability of the standard normal distribution function as follows:

$$Z_U = \begin{cases} 0 & \widehat{PNU}_{USL} \geq 0.5 \\ \Phi^{-1}\left(1 - \widehat{PNU}_{USL}\right) & 0.0 < \widehat{PNU}_{USL} < 0.5 \\ 4 & \widehat{PNU}_{USL} = 0 \end{cases} \tag{11}$$

where, $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

5) Finally, obtain the estimate of the process capability index ($\hat{C}_u$) of the concerned zero-inflated process as follows:

$$\hat{C}_u = (1/3) \times Z_U \tag{12}$$

If the value of index $\hat{C}_u$ is greater than 1, then the capability of the concerned zero-inflated process can be considered good. In this case, the process is capable of producing more than 99.865% conforming items, i.e. more than 99.865% of produced items will have nonconformities less than equal to $c_{USL}$ (the specified USL).

It is important to mention that if the value of $\widehat{PNU}_{USL}$ is more than or equal to 0.5, then $\hat{Z}_U$ (and hence $\hat{C}_u$) is considered as zero. When $\widehat{PNU}_{USL}$ is more than 0.5, it means that more than 50% of produced items are nonconforming, i.e. more than 50% of produced items will have nonconformities more than $c_{USL}$ (the specified USL). Thus, it is considered that the corresponding manufacturing process is not capable at all.

*4.1 Estimation of confidence interval of $C_u$*

Since $\hat{C}_u$ is a point estimate obtained from sample data, it is necessary to construct confidence interval (CI) of the capability index $C_u$ for inference purpose, especially when the sample size is relatively small. However, construction of CI using the sampling distribution of the estimated $\hat{C}_u$ is found to be quite difficult. Hence, we use Nagata and Nagahata (1994) proposed generalized approximation formula for construction of two-sided CI of $C_u$. According to Nagata and Nagahata (1994),

$$(1-\alpha)\% \text{ two-sided CI of } C_u = \left(\hat{C}_u - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{9n} + \frac{\hat{C}_u^2}{2(n-1)}}\ ,\ \hat{C}_u + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{9n} + \frac{\hat{C}_u^2}{2(n-1)}}\right) \tag{13}$$

where, $\alpha$ is the level of significance and (1- $\alpha$) is the confidence coefficient.

## 5. Analysis and Results

For the purpose of illustrations of computations of process capability indices using the proposed approach and assessing its effectiveness, two datasets published by past researchers are analyzed here as two case studies.

*5.1 Case study 1*

The yield, performance and reliability of semiconductor devices become more and more sensitive to particulate contamination as the chip density increases and semiconductor devices shrink in size. Therefore, monitoring of particles (particularly, which are bigger than a setting value) in the air of semiconductor manufacturing facilities is an important issue. The particles in the air can be measured and classified into discrete size using a laser particle counter.

With the aim to establish a control chart for monitoring of particles in a clean room of a semiconductor manufacturing facility, Tian et al. (2019) counted the particles greater than a setting value using a Lasair II-100 particle counter and collected 250 measurements on particle counts, which contain only 90 non-zero particle counts. This implies that these are zero-inflated count data. Therefore, it is decided to analyze the same data for evaluation of process capabilities using the proposed method and Patil and Shirke's (2012) approach. The full data set is available in Tian et al. (2019). The frequency distribution of different counts in Tian et al. (2019) observed data is shown in Table 1.

**Table 1.** Frequency distribution of different counts of particles

| Particle counts | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Observed frequency | 160 | 49 | 27 | 11 | 2 | 1 |
| Sample proportion | 0.640 | 0.196 | 0.108 | 0.044 | 0.008 | 0.004 |

The purpose of Tian et al.'s (2019) data collection was to establish appropriate control chart for monitoring the particles in a clean room of the semiconductor manufacturing facility and so they did not require taking into account the USL for the number of particles that can be tolerated in a sample unit. However, the USL must be known for computation of process capability index. So it is decided to assume that the USL for the number of particle is 5.

Now ZIP($\Omega, \lambda$) as well as ZINB($\Omega, k, \lambda$) distributions are fitted separately to these zero-inflated count data. Also the optimum log-likelihood value, AIC value and BIC value are observed for each case. These values are presented in Table 2.

**Table 2.** MLEs and values of model selection criterion for ZIP and ZINB distributions

| Elements | ZIP$(\Omega, \lambda)$ distribution | ZINB$(\Omega, k, \lambda)$ distribution |
|---|---|---|
| MLEs of the parameters | $\widehat{\Omega} = 0.53699$<br>$\hat{\lambda} = 1.10990$ | $\widehat{\Omega} = 0.88743$<br>$\hat{\lambda} = 0.67894$<br>$\hat{k} = 1.0$ |
| Optimum log-likelihood value | $-261.289$ | $-263.408$ |
| AIC value | 526.578 | 532.815 |
| BIC value | 533.621 | 543.380 |

It may be noted that AIC (or BIC) value reveals the relative goodness of two or more fitted models. But it does not say anything about the adequacy of the fitted model. Therefore, it is important to carry out the Chi-square goodness-of-fit for both the fitted distributions. The results of the Chi-square goodness-of-fit tests for the two fitted distributions are presented in Table 3.

**Table 3.** Chi-square goodness-of-fit tests for the fitted ZIP and ZINB distributions

| Particle counts | Observed frequency | Based on fitted ZIP distribution | | Based on fitted ZINB distribution | |
|---|---|---|---|---|---|
| | | Expected proportion | Expected frequency | Expected proportion | Expected frequency |
| 0 | 160 | 0.6400 | 160.00 | 0.64113 | 160.28 |
| 1 | 49 | 0.1964 | 49.11 | 0.21374 | 53.44 |
| 2 | 27 | 0.1090 | 27.25 | 0.08644 | 21.61 |
| 3 | 11 | 0.0403 | 10.08 | 0.03495 | 8.74 |
| 4 | 2 | 0.0112 | 2.80 | 0.01414 | 3.53 |
| 5 and above | 1 | 0.0031 | 0.76 | 0.00960 | 2.40 |
| Computed $\chi_v^2$ value | | $\chi_3^2 = 0.391$ | | $\chi_2^2 = 3.781$ | |
| Significance level $(p)$ | | $p = 0.942$ | | $p = 0.151$ | |

It can be observed from Table 3 that in both cases the computed Chi-square value is not statistically significant at 5% level. This implies that both the fitted models may be considered adequate to describe the particle counts data. However, the computed AIC as well as BIC values (see Table 2) are observed lower for the fitted ZIP distribution than the fitted ZINB distribution. Thus the ZIP model is selected for describing the particle count data, and subsequent analysis.

*Computation of process capability index*

The expected proportion of nonconforming items with respect to USL (i.e. *PNU*) is estimated using the fitted ZIP distribution as

$$\widehat{PNU} = 1 - \left[(1 - 0.53699) + 0.53699 \times \exp(-1.1099) + \sum_{y=1}^{5} 0.53699 \times \frac{\exp(-1.1099) \times 1.1099^y}{y!}\right]$$
$$= 0.00054$$

So the mapped Z-score in standard normal distribution is found as $Z_U = \Phi^{-1}(1 - 0.00054)$ =3.267. Thus, the process capability index, $C_u$ is estimated as $\hat{C}_u = (1/3) \times 3.267$ = 1.089, and using Nagata and Nagahata's (1994) generalized approximation formula, the 95% CI of $C_u$ is obtained as [0.9847, 1.1931].

It is of interest to estimate Patil and Shirke (2012) proposed $C_{pcu}^Z$index from the same dataset (2019) and compare the same with the estimated $\hat{C}_u$ value. The $C_{pcu}^Z$ index is a modified version of Perakis and Xekalaki (2005) proposed $C_{pcu}$ index and it is defined as follows:

$$C_{PCU}^Z = \frac{1-p_0}{\pi(1-p)} \qquad (15)$$

where $p_0$ is the desired proportion of conforming units with respect to *USL*, $p$ is the actual proportion of conforming units with respect to *USL*, and π (0<π<1) is the amount of inflation of zeroes in the count data. The recommend value for $p_0$ is 0.9973 (Perakis and Xekalaki, 2005), and here, $\hat{\pi} = \widehat{\Omega} = 0.53699$, $\hat{p} = 1 - \widehat{PNU} = 0.99946$. Thus, the estimate of $C_{PCU}^Z$ is obtained as $\hat{C}_{PCU}^Z = 9.245$, and using Nagata and Nagahata's (1994) generalized approximation formula, the 95% CI of $C_{PCU}^Z$ is derived as [8.432,10.058]. The estimates of $C_u$ and $C_{PCU}^Z$ indices and their 95% CI are presented in Table 4 for an easy comparison.

**Table 4.** Estimates of $C_u$ and $C_{PCU}^Z$ indices and their 95% CIs

| Sr. No. | Process capability index | Estimate of the index | 95% CI |
|---|---|---|---|
| 1 | $C_u$ | $\hat{C}_u = 1.089$ | (0.9847, 1.1931) |
| 2 | $C_{PCU}^Z$ | $\hat{C}_{PCU}^Z = 9.245$ | (8.432, 10.058) |

It may be worth to mention that since introduction of the concept of process capability index, conventionally the indices $C_p$, $C_{pu}$ and $C_{pl}$ are estimated from normal processes to facilitate better decision making in product and process management. The analysts/users of the indices can easily assess/predict the expected proportion of conforming products in the process outputs based on an estimated index value by virtue of the following relationships:

i) $P(LSL \leq X \leq USL) = 2\Phi\left(3 \times \hat{C}_p\right) - 1,$

ii) $P(X \leq USL) = P\left(z \leq 3 \times \frac{USL - \hat{\mu}}{3\hat{\sigma}}\right) = \Phi\left(3 \times \hat{C}_{pu}\right)$

iii) $P(X \geq LSL) = 1 - P(X < LSL) = 1 - \Phi\left(-3 \times \hat{C}_{pl}\right)$

For example, $\hat{C}_{pu} = 0.5$ implies that the process is capable of producing 93.319% conforming products with respect to $USL$, $\hat{C}_{pu} = 1$ implies that the process is capable of producing 99.865% conforming products and $\hat{C}_{pu} = 1.3$ implies that the process is capable of producing 99.995% conforming products. Over the years, process managers, engineers and other decision makers have become accustomed to relate the estimates of process capability indices and the expected proportion of product conformance to specifications in this way. Accordingly, general thumb rule is being followed among the users of the indices that the capability of a process is good if $\hat{C}_{pu} \geq 1$ and the capability is very good if $\hat{C}_{pu} \geq 1.33$.

Here the results shown in Table 4 reveal that the estimated values of the two indices obtained from the same dataset differs widely. The estimated $\hat{C}_u$ value ($= 1.089$) gives an idea that the concerned semiconductor manufacturing process is capable enough to maintain particle counts in the clean room within the USL. However, the estimated $\hat{C}_{PCU}^Z$ value ($= 9.245$) gives an impression that the concerned semiconductor manufacturing process is very highly capable (may be zero defective) for maintaining particle counts in the clean room within the USL. This is obviously a false impression because in this process, the expected proportion of nonconforming cases with respect to USL is found to be 0.00054 (540 ppm), which is not very low in reference to the concept of six sigma process.

*5.2 Case Study 2*

In the context of highlighting the problem of monitoring and control of a type of process in which long series with no nonconformities are observed together with occasional samples containing a large number of nonconformities, Xie and Goh (1993) presented a set of real life data on read-write errors discovered in 208 computer hard disks. It was observed that 180 hard disks had no error and only 28 hard disks contained non-zero errors. This implies that these are zero-inflated count data. Therefore, it is decided to analyze the same data for evaluation of process capabilities using the proposed method and Patil and Shirke's (2012) approach. The full data set is available in Xie and Goh (1993). The frequency distribution of different counts in Xie and Goh (1993) observed data is shown in Table 5.

**Table 5.** Frequency distribution of different counts of errors

| Counts of errors | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 9 | 11 | 15 and above |
|---|---|---|---|---|---|---|---|---|---|---|
| Observed frequency | 180 | 11 | 5 | 2 | 1 | 1 | 2 | 2 | 1 | 3 |
| Sample proportion | 0.865 | 0.053 | 0.024 | 0.010 | 0.005 | 0.005 | 0.010 | 0.010 | 0.005 | 0.014 |

Xie and Goh (1993) presented the count data in the context of process control and so they did not require taking into account the USL of the number of errors that can be tolerated in a hard disk. However, the USL must be known for computation of process capability index. So it is decided to consider that the USL for the number of errors in a hard disk is 10.

Now $\mathrm{ZIP}(\Omega, \lambda)$ as well as $\mathrm{ZINB}(\Omega, k, \lambda)$ distributions are fitted separately to these zero-inflated count data. Also the optimum log-likelihood value, AIC value and BIC value are observed for each case. These values are presented in Table 6.

**Table 6.** MLEs and values of model selection criterion for ZIP and ZINB distributions

| Elements | $\mathrm{ZIP}(\Omega, \lambda)$ distribution | $\mathrm{ZINB}(\Omega, k, \lambda)$ distribution |
|---|---|---|
| MLEs of the parameters | $\hat{\Omega} = 0.1346$<br>$\hat{\lambda} = 8.6413$ | $\hat{\Omega} = 0.21999$<br>$\hat{\lambda} = 1.69911$<br>$\hat{k} = 1.09145$ |
| Optimum log-likelihood value | $-192.985$ | $-154.54$ |
| AIC value | $389.969$ | $315.08$ |
| BIC value | $396.644$ | $325.09$ |

It may be noted that AIC (or BIC) value reveals the relative goodness of two or more fitted models. But it does not say anything about the adequacy of the fitted model. Therefore, it is important to carry out the Chi-square goodness-of-fit tests for both the fitted distributions. The results of the Chi-square goodness-of-fit tests for the two fitted distributions are presented in Table 7.

**Table 7.** Chi-square goodness-of-fit tests for the fitted ZIP and ZINB distributions

| Particle counts | Observed frequency | Based on fitted ZIP distribution | | Based on fitted ZINB distribution | |
|---|---|---|---|---|---|
| | | Expected proportion | Expected frequency | Expected proportion | Expected frequency |
| 0 | 180 | 0.8654 | 180.00 | 0.8590 | 178.67 |
| 1 | 11 | 0.0002 | 0.04 | 0.0481 | 10.00 |
| 2 | 5 | 0.0009 | 0.18 | 0.0293 | 6.09 |
| 3 | 2 | 0.0026 | 0.53 | 0.0178 | 3.71 |
| 4 | 1 | 0.0055 | 1.15 | 0.0109 | 2.26 |
| 5 | 1 | 0.0096 | 1.99 | 0.0066 | 1.37 |
| 6 | 2 | 0.0138 | 2.86 | 0.0040 | 0.84 |
| 9 | 2 | 0.0176 | 3.66 | 0.0009 | 0.19 |
| 11 | 1 | 0.0120 | 2.49 | 0.0003 | 0.07 |
| 15 and above | 3 | 0.0220 | 4.58 | 0.0186 | 3.87 |
| Computed $\chi_v^2$ value | | $\chi_7^2 = 2946.92$ | | $\chi_2^2 = 1.114$ | |
| Significance level ($p$) | | $p < 0.0001$ | | $p = 0.573$ | |

It can be observed from Table 7 that the computed Chi-square value for the fitted ZIP distribution is statistically significant at 5% level. This implies that the ZIP distribution is not a good fit for the sample dataset. However, the computed Chi-square value for the fitted ZINB distribution is observed not to be statistically significant at 5% level. This implies that the ZINB distribution adequately fits the sample data. The presence of over dispersion in the sample data caused by unobserved heterogeneity may be the possible cause for failure of the ZIP model to fit adequately. Since the ZINB model fit adequately, it is decided to use ZINB model for describing the error count data, and subsequent analysis.

*Computation of process capability index*

The expected proportion of nonconforming items with respect to USL (i.e. *PNU*) is estimated using the fitted ZINB distribution as

$$\widehat{PNU} = 1 - \left[(1 - 0.2199) + 0.2199\left(\frac{1.092}{2.791}\right)^{1.092} + \sum_{y=1}^{10} 0.2199 \times \frac{\Gamma(y+1.092)}{y!\Gamma(1.092)}\left(\frac{1.699}{2.791}\right)^{y}\left(\frac{1.092}{2.795}\right)\left(\frac{1.092}{2.795}\right)^{1.092}\right]$$
$$= 0.0190$$

So the mapped Z-score in standard normal distribution is found as $Z_U = \Phi^{-1}(1 - 0.0190) = 2.076$. Thus, the process capability index, $C_u$ is estimated as $\hat{C}_u = (1/3) \times 2.076 = 0.692$, and using Nagata and Nagahata's (1994) generalized approximation formula, the 95% CI of $C_u$ is obtained as [0.611, 0.772].

Patil and Shirke (2012) proposed $C_{pcu}^Z$ index is estimated from the same dataset (Xie and Goh, 1993). Here, $\hat{\pi} = \hat{\Omega} = 0.21999$, $\hat{p} = 1 - \widehat{PNU} = 0.981$. Thus, the estimate of $C_{PCU}^Z$ is obtained as $\hat{C}_{PCU}^Z = 0.646$ and using Nagata and Nagahata's (1994) generalized approximation formula, the 95% CI of $C_{PCU}^Z$ is derived as [0.569, 0.723]. The estimates of $C_u$ and $C_{PCU}^Z$ indices and their 95% CI are presented in Table 8 for an easy comparison.

**Table 8.** Estimates of $C_u$ and $C_{PCU}^Z$ indices and their 95% CIs

| Sr. No. | Process capability index | Estimate of the index | 95% CI |
|---|---|---|---|
| 1 | $C_u$ | $\hat{C}_u = 0.692$ | [0.611, 0.772] |
| 2 | $C_{PCU}^Z$ | $\hat{C}_{PCU}^Z = 0.646$ | [0.569, 0.723] |

The estimated $\hat{C}_u$ value $(= 0.692)$ gives an idea that the concerned process is not capable enough to maintain read-write errors in the hard disks within the USL. It is also computed that about 1.90% items are expected to be nonconforming with respect to the USL of number of read-write errors. In the sample data, 1.92% items (4 out of 208) have more than 10 read-write errors. So it may be considered that the $\hat{C}_u$ value gives quite a good assessment about the process capability. On the other hand, the estimated $\hat{C}_{PCU}^Z$ value computed based on the fitted ZINB distribution is found to be $0.646$, which is reasonably close to the estimated $\hat{C}_u$ value $(= 0.692)$. This implies that, in this case, Patil and Shirke (2012) proposed $C_{PCU}^Z$ index gives a reasonably acceptable estimate of the true capability of the concerned zero-inflated process.

It may be worth to mention that Patil and Shirke (2012) analyzed the same dataset (Xie and Goh, 1993) for illustrating their proposed method for estimating capability of a zero-inflated process. The estimated $\hat{C}_{PCU}^Z$ value obtained by them was 0.0549, which was much smaller than the estimated $\hat{C}_{PCU}^Z$ value obtained by us. The main reason behind this difference is that they carried out all the computation based on fitting of ZIP distribution to the observed data. But actually ZINB distribution fit well to the observed data, not ZIP distribution, which is evident in the results presented in Table 7. This difference in the estimated values $\hat{C}_{PCU}^Z$ also highlights the importance of fitting appropriate zero-inflated distribution.

It may be noted that Xie et al. (2001) have also used the same dataset (Xie and Goh, 1993) for illustration of their proposed control chart for ZIP processes. They applied various tests of Poisson distribution and zero-inflated Poisson (ZIP) alternative to the dataset, and all the tests suggested that the ZIP model should be used instead of the conventional Poisson model. Accordingly, they fitted ZIP model to the dataset and determined the upper control limit of the proposed control chart. However, they did not consider fitting of zero-inflated negative binomial (ZINB) distribution. On the other hand, in the current research works, both ZIP and ZINB distributions are fitted separately to the same dataset (Xie and Goh, 1993). Also the optimum log-likelihood value, AIC value and BIC value are observed in each case for identification of better fitted distribution. Based on examination of the log-likelihood, AIC and BIC values, it is determined that ZINB model fit better to the dataset (Xie and Goh, 1993). Accordingly, expected proportion of nonconforming items with respect to USL (i.e. *PNU*) is estimated using the fitted ZINB distribution and process capability index is computed using the proposed method. The analysis of the current research is indicative that the performance of the proposed control chart of Xie et al. (2001) could be better if they would have fitted ZINB distribution to the dataset and determined the upper control limit based on the fitted ZINB distribution.

## 6. Conclusions

One of the important characteristics of zero-inflated processes is that these have more count of zeros than are expected under chance variation of its underlying Poisson or other count distribution, and therefore, the standard Poisson or other count distribution fails to model sample count data obtained from the zero-inflated processes. These processes are commonly modelled by ZIP or ZINB distribution. Evaluation of capabilities of zero-inflated processes in producing outputs within specification limit(s) is an important issue. For the purpose of assessing capability of a zero-inflated process, estimation of $C_{PCU}^Z$ index is proposed in literature. But, it does not always represent the true capabilities of zero-inflated processes, and sometimes it gives very misleading impression about the capability of the concerned process (as observed in case study 1). In this article, the concept of Borges and Ho (2001) is applied to zero-inflated processes and a new approach for computation of process capability index of zero-inflated processes is developed. The proposed method reveals the true capabilities of zero-inflated processes consistently. The results of analysis of two datasets (published in literature) validate the same. Thus, the proposed method can be regarded as a procedure for evaluating capability of zero-inflated processes.

One important assumption in this study is that only a single type of random shock occurs in the zero-inflated process. But in real life zero-inflated processes, multiple types of shocks may occur resulting in different types of nonconformities. In such situations multivariate ZIP or multivariate ZINB models may be fitted to the sample dataset. Future research need to be carried out for evaluating capability of multivariate zero-inflated processes.

**Acknowledgement**
We are sincerely grateful to the referee for his/her valuable comments which have helped in improving the overall clarity and readability of this article.

## References


Alevizakos, A. and Koukouvinos, C., 2020. Monitoring of zero-inflated binomial processes with a DEWMA control chart. *Journal of Applied Statistics.* https://doi.org/10.1080/02664763.2020. 1761950.

Borges, W. and Ho, L. L., 2001. A fraction defective based capability index. *Quality and Reliability Engineering International*, Vol. 17, pp. 447-458. https://doi.org/10.1002/qre.438

Chaney, S. E. P., Charity, M., David, M. and Aban, I., 2013. Zero-inflated and overdispersed: what's one to do? *Journal of Statistical Computation and Simulation*, Vol. 83, pp. 1671–1683. https://doi.org/10.1080/00949655.2012.668550

Chang, T. C. and Gan, F. F., 1999. Charting techniques for monitoring a random shock process. *Quality and Reliability Engineering International*, Vol. 15, pp. 295–301.

Chen, K. S., Wang, K. J. and Chang, T. C., 2017. A novel approach to deriving the lower confidence limit of indices $C_{pu}$, $C_{pl}$, and $C_{pk}$ in assessing process capability. *International Journal of Production Research*, Vol. 55, pp. 4963-4981. https://doi.org/10.1080/00207543.2017.1282644

Clements, J. A., 1989. Process capability calculations for non-normal distributions. *Quality Progress*, Vol. 22, pp. 95-100.

Dodge, H. F., 1943. A sampling plan for continuous production. *Annals of Mathematical Statistics*, Vol. 14, pp. 264–279.

Favero, L. P., Hair, J. F., Jr., Souza, R.D.F., Albergaria, M. and Brugni, T.V., 2021. Zero-inflated generalized linear mixed models: a better way to understand data relationships. *Mathematics*, Vol. 9, pp. 1-28. https://doi.org/10.3390/math9101100

Gupta, P. L., Gupta, R. C. and Tripathi, R. C., 1996. Analysis of zero-adjusted count data. *Computational Statistics & Data Analysis*, Vol. 23, pp. 207–218. https://doi.org/10.1016/S0167-9473(96)00032-1

Kane, V. E., 1986. Process capability indices. *Journal of Quality Technology*, Vol. 18, pp. 41-52. https://doi.org/10.1080/00224065.1986.11978984

Kotz, S. and Johnson, N. L., 2002. Process capability indices - a review, 1992-2000. *Journal of Quality Technology*, Vol. 34, pp 2-19. https://doi.org/10.1080/00224065.2002.11980119

Kumar, C. S. and Ramachandran, R., 2020. On some aspects of a zero-inflated overdispersed model and its applications. *Journal of Applied Statistics*, Vol. 47, pp. 506-523. https://doi.org/10.1080/02664763.2019.1645098

Lambert, D., 1992. Zero–inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, Vol. 34, pp. 1-14. https://doi.org/10.2307/1269547

Loganathan, A. and Shalini, K., 2014. Determination of single sampling plans by attributes under the conditions of zero□inflated Poisson distribution. *Communications in Statistics – Simulation and Computation*, Vol. 43, pp. 538□548. https://doi.org/10.1080/03610918.2012.707723

Mahmood, T. and Xie, M., 2019. Models and monitoring of zero-inflated processes: The past and current trends. *Quality and Reliability Engineering International*, Vol. 35, pp. 2540-2557. https://doi.org/10.1002/qre.2547

Maiti, S. S., Saha, M. and Nanda, A. K., 2010. On generalising process capability indices. *Journal of Quality Technology and Quantitative Management*, Vol. 7, pp. 279-300. https://doi.org/10.1080/16843703.2010.11673233

Martin, J. and Hall, D. B., 2017. Marginal zero□inflated regression models for count data. *Journal of Applied Statistics*, Vol. 44, pp. 1807□1826. https://doi.org/10.1080/02664763.2016.1225018

Montgomery, D. C., 2012. *Introduction to Statistical Quality Control*, 7$^{th}$ ed. Wiley, New York, 2012.

Nagata, Y. and Nagahata, H., 1994. Approximation formulas for the confidence intervals of process capability indices. *Okayama Economic Review*, Vol. 25, pp. 301-314.

Pal, S. and Gauri, S. K., 2020. Measuring capability of a binomial process. *International Journal of Engineering, Science and Technology*, Vol. 12, pp. 25-37. https://doi.org/10.4314/ijest.v12i1.3

Pal, S. and Gauri, S. K., 2020. Measuring capability of a Poisson process: Relative goodness of the estimates by different approaches. *International Journal of Engineering, Science and Technology*, Vol. 12, pp. 1-13. https://doi.org/10.4314/ijest.v12i4.1

Patil, M. K. and Shirke, D.T., 2012. Process capability index for zero-inflated Poisson process. *International Journal of Mathematical Sciences & Engineering Applications*, Vol. 6, pp. 87-97.

Pearn, W. L. and Chen, K. S., 1995. Estimating process capability indices for non-normal Pearsonian populations. *Quality and Reliability Engineering International*, Vol. 11, pp. 389-391. https://doi.org/10.1002/qre.4680110510

Perakis, M. and Xekalaki, E., 2005. A process capability index for discrete processes. *Journal of Statistical Computation and Simulation*, Vol. 75, pp. 175-187. https://doi.org/10.1080/00949650410001687244

Phang, Y. N. and Loh, E. F., 2013. Zero inflated models for overdispersed count data, *International Journal of Health and Medical Engineering*, Vol. 7, pp. 1331-1333.

Polhemus, N. W., 2018. *Process Capability Analysis: Estimating Quality*. CRC press, Taylor & Francis Group, Boca Ration, Florida.

Rakitzis, A. C., Maravelakis, P. E. and Castagliola, P., 2016. CUSUM control charts for the monitoring of zero□inflated binomial processes. *Quality and Reliability Engineering International*, Vol. 32, pp. 465□483. https://doi.org/10.1002/qre.1764

Rao, G. S. and Aslam, M., 2017. Resubmitted lots with single sampling plans by attributes under the conditions of zero□inflated Poisson distribution. *Communications in Statistics – Simulations and Computation*, Vol. 46, pp. 1814□1824. https://doi.org/10.1080/03610918.2015.1016236

Sim, C. H. and Lim, M. H., 2008. Attribute charts for zero-inflated processes, *Communications in Statistics – Simulations and Computation*, Vol. 37, pp.1440-1452. https://doi.org/10.1080/03610910801983145

Tian, W., You, H., Zhang, C., Kang, S., Jia, X. and Chien, W.K., 2019. Statistical process control for monitoring the particles with excess zero counts in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, Vol. 32, pp.93□103. https://doi.org/10.1109/TSM.2018.2882862

Wagh, Y. S. and Kamalja, K. K., 2018. Zero-inflated models and estimation in zero-inflated Poisson distribution. *Communications in Statistics – Simulations and Computation*, Vol. 47, pp. 2248-2265. https://doi.org/10.1080/03610918.2017.1341526

Workie, M. S. and Azene, A. G., 2021. Bayesian zero-inflated regression model with application to under-five child mortality. *Journal of Big Data*, Vol. 8, pp. 1-24. https://doi.org/10.1186/s40537-020-00389-4

Xie, M. and Goh, T., 1993. SPC of a near zero-defect process subject to random shocks. *Quality and Reliability Engineering International*, Vol. 9, pp. 89-93.

Xie, M. He, B. and Goh, T. N., 2001. Zero-inflated Poisson model in statistical process control. *Computational Statistics & Data Analysis*, Vol. 38, pp. 191-201. https://doi.org/10.1016/S0167-9473(01)00033-0

Yang, J., Xie, M. and Goh, T. N., 2011. Outlier identification and robust parameter estimation in a zero-inflated Poisson model. *Journal of Applied Statistics*, Vol. 38, pp. 421-430. https://doi.org/10.1080/02664760903456426

Yeh, F. B. and Bhattacharya, S., 1998. A robust process capability index, *Communications in Statistics - Simulation and Computation*, Vol. 27, pp. 565-589. https://doi.org/10.1080/03610919808813495

Zeileis, A., Kleiber, C. and Jackman, S., 2008. Regression models for count data in R. *Journal of Statistical Software*, Vol. 27, pp. 1-25. https://doi.org/10.18637/jss.v027.i08

Zhao, Y., Lee, A. H., Burke, V. and Yau, K. K.W., 2009. Testing for zero-inflation in count series: application to occupational health. *Journal of Applied Statistics*, Vol. 36, pp. 1353-1359. https://doi.org/10.1080/02664760802653586

**Biographical notes**

**Dr. Surajit Pal** is a Faculty Member in the Statistical Quality Control and Operations Research Unit of the Indian Statistical Institute, Chennai Centre, India. His fields of interest are quality engineering, process optimization, statistical quality control and multiple response optimization. He has published about forty five papers in different National and International journals.

**Dr. Susanta Kumar Gauri** is a Faculty Member in the Statistical Quality Control and Operations Research Unit of the Indian Statistical Institute, Kolkata, India. His fields of interest are quality engineering, process optimization, statistical quality control and multiple response optimization. He has published about sixty papers in different National and International journals.