Chapter 6

Zero-Inflated Models in Statistical Process Control

6.0 Introduction

In statistical process control Poisson distribution and binomial distribution play important role. There are situations wherein the process leads to zero-inflation in its output. Traditional process in monitoring manufacturing and biological studies usually involve selecting random samples at regular intervals. Due to technological advancement and automation of manufacturing processes, a well-designed process could have more count of zeros than are expected under chance variation of its underlying Poisson or binomial distribution. One such process is the thermo sonic wire bonding process of an integrated circuit assembly (Chang and Gan 2001). The bonding machine has a closed loop controlled system to detect and rectify any non-conformity generated during the bonding process.

The excess number of zeros in binomial count can also be found in the biological control of pests. The Shewhart c-chart and np-chart are widely used to monitor processes with Poisson and binomial counts. However, as both the Poisson and binomial distributions tend to underestimate the mean and variability of the zero-inflated count, the resulting attribute charts have tighter control limits which subsequently lead to a higher false alarm rate in detecting out-of-control signals.

When the model is used in statistical process control, control limits can then be derived based on the zero-inflated Poisson model and zero-inflated binomial model. The classical Shewhart c-chart and np-chart constructed based on the Poisson and binomial distributions are inappropriate in monitoring zero-inflated counts. They tend to underestimate the dispersion of zero-inflated counts and subsequently lead to higher false alarm rate in detecting out-of-control signals (Sim and Lim (2008)).

Another drawback of these charts is that their 3-sigma control limits, evaluated based on the asymptotic normality assumption of the attribute counts; fail to provide the required false alarm rate for the parameter under study. To overcome these drawbacks, Sim and Lim (2008) proposed the zero-inflated models which will take the account of the excess number of zeros. They have provided attribute charts for zero-inflated processes and estimated the Poisson parameter θ from zero-inflated Poisson model. Further they have used the model to construct a one sided c - chart with its upper control limit constructed based on the Jeffreys prior interval or Blyth - Still interval of the binomial proportion θ. However in the study of performance of the one-sided c chart using average run length (ARL) criterion Sim and Lim (2008) have not used ZIP model. ARL values are evaluated using classical Poisson model. Additionally, a simple two-of-two control rule is also recommended to improve further on the performance of these two proposed charts. Xie et al. (2001) proposed ZIP model instead of conventional Poisson model in statistical process control.

In the following the zero-inflated models for attribute counts with excess number of zeros are discussed.

6.1 Control Charts for Poisson and Binomial Processes

In any production process there exist certain amounts of inherent or natural variability. This natural variability or "background noise" is the cumulative effect of many small, essentially unavoidable causes. In the frame work of statistical quality control, this natural variability is often called a stable system of chance causes. A process that is operating with only chance causes of variation alone is said to be in statistical control. In other words, the chance causes are an inherent part of the process. The other kind of variability may be present in the process output is an unnatural variation which is known as assignable causes of variation. Such variability is generally large when compared to the back ground noise. A process operating in the presence of assignable causes is said to be out of control.

In recent years due to advanced technologies there is a competitive environment in manufacturing industries. Therefore, manufacturer has to produce products of very high quality. However, it is not an easy task to maintain such a high quality of product. The main reason behind this is the inherent variability in the quality of the product. This product may be components require for automobiles, various equipments like power generating sets, electric motor etc. If the variability can only be described in statistical terms, the statistical methods play an important role in the quality improvement. The technique comprising these statistical methods is named as 'Statistical Quality Control' (SQC). The SQC techniques are used for the process control, the product control and to design scientific experiments.

A General Model for Control Charts:

Let W be a statistic that measures a quality characteristic of interest. Suppose that the mean of W is μ_W and the standard deviation of W is σ_W. The control limits become

$$UCL = \mu_W + L\sigma_W$$

$$CL = \mu_W$$

$$LCL = \mu_W - L\sigma_W , \qquad \ldots(6.1.1)$$

where L is the distance of the control limits from the centre line, expressed in standard deviation units. It is customary to choose $L = 3$. The control limits taking $L = 3$ are called Shewhart's control limits. The concept of control charts was first proposed by Dr. Walter A. Shewhart and control chart developed according to these principles are often called Shewhart control charts.

The Classical $C-$Control Chart for Poisson Count

Poisson distribution is a natural choice for modeling the occurrence of random event or the number of nonconformities (X) found in a subgroup of size n. The classical Shewhart c-chart has been widely used to monitor processes with Poisson count. The 3-sigma control limits of a classical two-sided c-chart when standards are known, in monitoring a Poisson process with mean θ are given by

$$UCL_c = \theta + z_{\alpha/2}\sqrt{\theta}$$

$$CL_c = \theta$$

$$LCL_c = \max(0, \theta - z_{\alpha/2}\sqrt{\theta}) \qquad \ldots(6.1.2)$$

When standards are unknown the parameter θ is replaced by $\overline{X}$ (i.e. mean), where $\overline{X} = \sum_{i=1}^{n} \frac{x_i}{n}$ and the corresponding control limits are

$$UCL_c = \overline{X} + z_{\alpha/2}\sqrt{\overline{X}}$$

$$CL_c = \overline{X}$$

$$LCL_c = \max(0, \overline{X} - z_{\alpha/2}\sqrt{\overline{X}}) \qquad ...(6.1.3)$$

The np-Control Chart

The np control chart has been widely used in industry to monitor the number of nonconforming units in sample of n units inspected. A nonconforming unit (d) is the number of units that fails to meet at least one specified requirement in a subgroup of size n. The control limits using binomial distribution are as follows

$$UCL = np + 3\sqrt{np(1-p)}$$
$$CL = np$$
$$LCL = np - 3\sqrt{np(1-p)} \qquad \ldots(6.1.4)$$

Suppose, we have m subgroups samples each of size n. And if a standard value of p is unknown, then $\overline{p}$ can be used to estimate p, where $\overline{p} = \frac{\sum_{i=1}^{m} d_i}{mn}$.

Then the centre line and control limits are as follows:

$$UCL = n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})}$$
$$CL = n\overline{p}$$
$$LCL = n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})} \qquad \ldots(6.1.5)$$

Now, we introduce the concept of run-length, average run-length and expression for average run-length.

Run Length

The run-length of control chart is the number of samples required to detect the out-of-control state. Clearly run-length is a positive integer valued random variable with support 1, 2, 3….

Average Run Length (ARL)

Average run length is defined as the average number of points that must be plotted before a point indicates an out of control condition. The performances of the control chart are often described in terms of their run length distribution. When the process is in-control we expect that ARL should be as large as possible and when process goes out-of-control, ARL should be as small as possible. Comparison of one control chart with the other is usually done through ARL.

One-of-One **Control Rule**

In this rule, a process is declared to be out of control if single point falls outside the control limits.

Two-of-Two C**ontrol Rule**

In this rule, a process is declared to be out of control if two successive points fall outside the control limits. It is found that two-of-two control rule are more sensitive than one-of-one control rules, in the sense that, the former detect out of control state early.

6.2 Control Chart for Zero-Inflated Poisson Count

The excess number of zero counts causes reduction in the value of mean. Consequently, this reduces the natural variability within control limits. As a result of this, there will be many false alarms. Therefore, we implement zero-inflated models to construct control limits. Further to assess the performance of the proposed control charts we have used confidence interval that provides good coverage probability.

First, consider the control limits based on ZIP model. Note that the Poisson parameter θ is replaced by its maximum likelihood estimator $(\hat{\theta})$. The control limits based on this mle are given by,

$$UCL_c = \hat{\theta} + z_{\alpha/2}\sqrt{\hat{\theta}}$$

$$CL_c = \hat{\theta}$$

$$LCL_c = \max(0, \hat{\theta} - z_{\alpha/2}\sqrt{\hat{\theta}}) \qquad \ldots(6.2.1)$$

Control Chart Based On *Two- Of- Two* Control Rule

To improve further the performance of the control chart a simple *two-of- two* control rule is used and the control limits are set so that rule gives desired in control ARL. These limits are obtained by using simulation. R program has been developed for the same (Appendix 3).

The Control Chart for Zero- Inflated Binomial Count

When an excess numbers of zero counts are present then the control limits are evaluated based on a process fraction nonconforming p estimated from ZIB model. This p is denoted by $\hat{p}$ and it is the solution of the following equation,

$$\sum_{i=1}^{n} x_i - \frac{n\,\hat{p}(n-n_0)}{(1-(1-\hat{p})^n} = 0 .$$

The control limits then are given by

$$UCL_{np} = n\hat{p} + z_{\alpha/2}\sqrt{n\hat{p}\,(1-\hat{p})}$$

$$CL_{np} = n\hat{p}$$

$$LCL_{np} = n\hat{p} - z_{\alpha/2}\sqrt{n\hat{p}\,(1-\hat{p})} \qquad \ldots(6.2.2)$$

Performance of the ZIP Control Chart

Now, we shall consider the performance of the one-sided c- chart using its average run length ARL(Δ). It denotes the number of inspection units required to inspect until one is able to detect an upward shift of the Poisson process mean from θ_0 to $\theta_1 = \theta_0 + \Delta$, $\Delta \geq 0$. The in-control average run length ARL ($\Delta = 0$), is denoted by ARL$_0$. It is the average number of inspection units required to inspect before the c- chart falsely detect an out-of-control signal when the process is actually in-control. The out of control average run length ARL$(\Delta \neq 0)$, namely ARL$_1$, is the average number of inspection units that should be inspected to correctly detect an out-of-control signal when the process is actually out-of-control. In general we require a reasonably large ARL$_0$ value in order to have a low false alarm rate, and small ARL$_1$ values to enable rapid detection of shifts in process mean. The average run length of the c chart is given by

$$\text{ARL } (\Delta) = \frac{1}{P_c(\Delta)},$$

in which $P_c(\Delta) = p(X > UCL_c(\theta_0) \mid \theta = \theta_1)$

$$= \sum_{x=UCL_c(\theta_0)+1}^{\infty} \frac{e^{-\theta_1}\theta_1^{\,x}}{x!}. \qquad \ldots(6.2.3)$$

$P_c(\Delta)$ is the probability that the Poisson count of a given inspection unit is plotted above the control limit $UCL_c(\theta_0)$. Sim and Lim (2008) reported ARL values for zero-inflated Poisson process. However it is found that ARL calculation are based on classical Poisson model and hence are incorrect.

The performance of the classical c chart constructed with standards known, standards unknown, ZIP model and two-of two control rule in monitoring ZIP process with parameters $\theta = 4.0$ (0.5) 5.5 is given in Table 6.2.1. Examination of Table 6.2.1 reveals that for a process with zero-inflated parameter $\pi = 0.6$, and Poisson parameter $\theta = 4.0$; (i) the classical c-chart (standards unknown), constructed without taking in to account of ZIP count, yields an unacceptably small ARL_0 of 15.0215 ; (ii) the c-chart constructed based on the ZIP model yields a bigger ARL_0 of 205.2845; (iii) the classical c–chart (standards known) yields 203.5522; and (iv) the c-chart constructed based on *two-of- two* control rule yields a larger ARL_0 value of 243.8718.

For the case $\theta = 4.5$, (i) the classical c-chart (standards unknown), constructed without taking in to account of ZIP count, yields small ARL_0 of 19.2018; (ii) the c-chart constructed based on the ZIP model yields much larger ARL_0 of 691.1483; (iii) the classical c–chart (standards known) yields 254.3254; and (iv) the c-chart constructed based on *two-of- two* control rule yields a ARL_0 value of 393.5261.

The performance of the classical c chart with inflation parameter $\pi = 0.4$ and Poisson parameter $\theta = 4.0$ (0.5) 5.5 are given in Table 6.2.2.

Table : 6.2.1

ARL performance of attribute charts in monitoring ZIP processes

($\pi = 0.6$, and $\theta = 4.0$ (0.5) 5.5)

		ARL of c chart with						
θ	θ_1	standards known	standards unknown	ZIP model	two-of-two control rule-1	Adjusted ARL of c chart ZIP	Adjusted ARL for two-of-two chart	Actual ARL for ZIP model
	UCL	9	5	9	6	9	6	
4.00	4.00	**203.5522**	**15.021**	**205.2845**	**243.8718**	203.5522	203.5522	204.9455
	4.80	66.3461	7.9835	68.2963	72.3945	67.7200	60.4254	66.2923
	5.60	28.6513	5.1072	28.0752	31.5682	27.8383	26.3490	28.1865
	6.40	14.6707	3.6265	14.6485	17.9370	14.5249	14.9715	14.5941
	UCL	10	7	11	7	11	7	
4.50	4.50	**254.3254**	**19.2018**	**691.1483**	**393.5261**	254.3254	254.3254	693.2072
	5.40	73.1962	9.4581	172.3299	99.1555	63.4131	64.0815	173.0409
	6.30	29.5977	5.6060	59.4104	38.1287	21.8616	24.6416	60.0943
	7.20	14.8110	3.8222	27.0266	19.6156	9.9451	12.6770	26.4946
	UCL	11	7	11	8	11	8	
5.0	5.00	306.0902	12.7055	308.2614	626.0450	306.0902	306.0902	305.6370
	6.00	83.0512	6.5070	82.5209	131.2260	81.9397	64.1599	82.9519
	7.00	31.6023	4.1094	31.5868	44.8271	31.3643	21.9172	31.2404
	8.00	14.7838	3.0933	15.1220	21.7713	15.0155	10.6446	14.8910
	UCL	12	7	12	8	12	8	
5.50	5.50	371.1953	8.7306	375.2733	264.7746	371.1953	371.1953	374.4576
	6.60	93.7234	4.7974	93.8611	65.5033	92.8411	91.8310	93.1684
	7.70	33.1591	3.2647	32.6317	26.2859	32.2771	36.8510	33.0512
	8.80	14.9822	2.5845	15.4726	14.7322	15.3045	20.6535	15.1292

Table: 6.2.2

ARL performance of attribute charts in monitoring ZIP processes

($\pi = 0.4$, and $\theta = 4.0$ (0.5) 5.5)

		ARL of c chart with				Adjusted ARL of c chart ZIP	Adjusted ARL for two-of-two chart	Actual ARL for ZIP model
θ	θ_1	standards known	standards unknown	ZIP model	two-of-two control rule-1			
	UCL	9	5	9	6	9	6	9
4.00	4.00	314.6101	11.6475	307.2295	534.6570	314.6101	314.6101	307.4183
	4.80	100.0760	7.2741	97.5526	156.1884	99.8961	91.9065	99.4385
	5.60	42.4400	5.1078	41.8111	65.4201	42.8155	38.4954	42.2797
	6.40	22.1013	3.9796	22.0251	36.0239	22.5542	21.1977	21.8912
	UCL	10	6	10	7	10	7	10
4.50	4.50	371.4469	14.7109	372.2706	850.3266	371.4469	371.4469	374.8872
	5.40	113.4831	8.3425	111.4406	208.5622	111.1940	91.1059	111.0432
	6.30	44.2597	5.6009	44.8062	80.0093	44.7071	34.9503	44.4211
	7.20	22.1137	4.2997	22.0198	40.4738	21.9711	17.6801	22.0607
	UCL	11	6	11	8	11	8	11
5.00	5.00	464.2578	10.4784	466.0903	1373.5670	464.2578	464.2578	458.4555
	6.00	125.4139	6.3106	124.0049	289.1661	123.5174	97.7365	124.4279
	7.00	47.8497	4.4961	46.4122	93.9766	46.2297	31.7636	46.8609
	8.00	22.2588	3.6309	22.226	44.9074	22.1386	15.1784	22.3365
	UCL	12	7	12	8	12	8	12
5.50	5.50	558.9356	13.2538	559.2587	583.7251	558.9356	558.9356	561.6864
	6.60	140.0044	7.2491	141.3947	137.9524	141.3130	132.0939	139.7527
	7.70	49.5208	4.9968	49.5658	53.8251	49.5372	51.5393	49.5768
	8.80	22.6833	3.8084	22.7722	29.1510	22.7590	27.9130	22.6938

Performance of the np Chart

The average run length of the np chart is detecting an upward shift of binomial parameter p (such as proportion of nonconforming) from p to $p_1 = p + \Delta, \quad \Delta > 0,$ is given as

$$\text{ARL } (\Delta) = \frac{1}{P_{np}(\Delta)}$$

in which $P_{np}(\Delta) = p(X > UCL_{np}(p_0) \mid p = p_1)$

$$= \sum_{x=UCL_c(p_0)+1}^{n} \binom{n}{x} p_1^{\,x} (1 - p_1)^{n-x} \qquad \ldots(6.2.4)$$

We have conducted a comprehensive study on the performance of various np charts. Table 6.2.3 reports only the ARLs of the np-chart with a *two- of -two* control rule in monitor ring the binomial processes with zero-inflated parameter $\pi =$ 0.6 and selected the classical np- charts constructed with and without using the ZIB model are given as well for comparison purposes. Table 6.2.3 shows that, for a ZIB model with parameters $\pi = 0.6$, $n =$ 653, and $p =$ 0.004 that result in a variance of 2.00(0.30)2.60: (i) the classical np-chart, (standards known) constructed without taking into account the excess number of zeros in the sample, yields ARL_0 of 300.650; (ii) the classical np-chart (standards unknown), constructed without taking into account the excess number of zeros in the sample, yields an undesirably small ARL_0 of 33.108; (iii) the ARL_0 of the np-chart increases to 321.406 when the ZIB model is used to fit the ZIB count; (iv) the np -chart yields ARL_0 value of 193.157 when two-of- two rule is applied; and (v) the actual ARL_0 is 310.273

The performance of the np chart with inflation parameter $\pi = 0.4$ and $p = 0.004\ (0.003)\ 0.010$ are given in Table 6.2.4

Table : 6.2.3

ARL performance of attribute charts in monitoring ZIB processes

($\pi = 0.6$, and $p = 0.004$ (0.003) 0.010)

p	p_1	Sample Size n / Variance	ARL of np - chart with				Adjusted ARL of np chart ZIB model	Adjusted ARL for two-of-two chart	Actual ARL for ZIB model
			standards known	standards unknown	ZIB model	two-of-two control rule			
0.004	**0.0040**	**502**	347.515	31.298	364.358	146.621	347.515	347.515	367.626
	0.0048	(2.00)	143.507	17.283	145.356	64.448	138.637	152.751	143.471
	0.0056		68.965	10.594	66.931	35.528	63.837	84.207	68.024
	0.0040	**577**	616.209	56.110	187.699	11.041	616.209	616.209	177.796
	0.0048	(2.30)	226.463	27.062	70.429	140.647	231.216	210.851	72.812
	0.0056		97.066	15.286	35.770	63.255	117.432	94.828	36.187
	0.0040	**653**	**300.650**	**33.108**	**321.406**	**193.157**	**300.650**	**300.650**	**310.273**
	0.0048	(2.60)	113.271	17.009	113.424	73.415	106.099	114.271	112.316
	0.0056		51.731	10.205	53.126	35.981	49.695	56.005	50.739
0.010	**0.0070**	**288**	362.040	31.719	369.104	146.673	362.040	362.040	365.709
	0.0084	(2.00)	141.529	17.288	146.962	63.628	144.149	157.057	142.554
	0.0098		67.344	10.831	66.978	34.573	65.696	85.339	67.521
	0.0070	**331**	671.267	19.705	172.702	397.599	671.267	671.267	176.460
	0.0084	(2.30)	231.128	11.136	69.635	134.901	270.661	227.753	72.201
	0.0098		95.357	7.414	35.630	62.990	138.489	106.346	35.860
	0.0070	**374**	310.105	33.232	314.313	192.751	310.105	310.105	310.709
	0.0084	(2.60)	117.557	17.149	115.313	73.141	113.709	117.672	112.281
	0.0098		48.439	10.146	50.624	36.309	49.920	58.415	50.648
0.010	**0.0010**	**202**	335.741	99.897	369.624	141.447	335.741	335.741	367.715
	0.0120	(2.00)	144.770	45.855	142.099	61.896	129.073	146.916	143.071
	0.0140		69.326	25.094	68.784	34.724	62.479	82.422	67.655
	0.0010	**232**	683.379	55.500	173.560	404.537	683.379	683.379	177.473
	0.0120	(2.30)	215.644	26.330	69.654	135.857	274.257	229.502	72.482
	0.0140		94.236	15.299	37.079	62.995	145.996	106.417	35.941
	0.0100	**263**	331.311	33.329	302.827	187.238	331.311	331.311	307.035
	0.0120	(2.60)	114.857	16.553	105.332	72.314	115.240	127.958	110.870
	0.0140		51.772	10.062	50.115	35.239	54.829	62.354	49.986

Table : 6.2.4

ARL performance of attribute charts in monitoring ZIB processes

$\pi = 0.4$, and $p = 0.004$ (0.003) 0.010.

p	p_1	Sample Size n / Variance	ARL of np - chart				Adjusted ARL of np chart ZIB	Adjusted ARL for two-of-two chart	Actual ARL for ZIB model
			with standards known	with standards unknown	with ZIB model	with two-of-two control rule			
0.004	**0.0040**	**502**	555.101	17.555	526.984	325.605	555.101	555.101	551.439
	0.0048	(2.00)	220.905	11.025	215.406	130.806	226.899	223.001	215.206
	0.0056		106.474	8.143	97.225	55.016	102.412	93.793	102.036
	0.0040	**577**	943.439	12.454	266.714	895.377	943.439	943.439	266.694
	0.0048	(2.30)	328.407	8.321	106.162	306.943	375.523	323.419	109.218
	0.0056		141.557	6.245	53.919	136.083	190.726	143.387	54.281
	0.0040	**653**	501.033	19.852	464.375	434.172	501.033	501.033	465.410
	0.0048	(2.60)	163.505	12.000	164.182	158.066	177.143	182.407	168.474
	0.0056		77.416	8.259	77.323	76.669	83.427	88.476	76.108
0.010	**0.0070**	**288**	573.553	17.226	559.469	318.283	573.553	573.553	548.564
	0.0084	(2.00)	221.714	11.036	213.406	134.468	218.778	242.313	213.830
	0.0098		99.710	8.079	103.471	72.510	106.076	130.665	101.282
	0.0070	**331**	943.723	12.203	271.798	890.233	943.723	943.723	264.689
	0.0084	(2.30)	324.680	8.179	109.698	298.694	380.888	316.641	108.302
	0.0098		146.105	6.129	57.324	134.437	199.037	142.515	53.790
	0.0070	**374**	458.037	20.010	455.706	427.213	458.037	458.037	466.063
	0.0084	(2.60)	176.080	11.912	166.902	156.873	167.756	168.192	168.421
	0.0098		75.914	8.184	73.428	76.105	73.804	81.596	75.973
0.010	**0.0010**	**202**	583.145	17.175	564.285	309.346	583.145	583.145	551.573
	0.0120	(2.00)	211.674	10.963	218.935	135.324	226.252	255.098	214.606
	0.0140		102.713	7.808	95.702	72.154	98.900	136.017	101.481
	0.0010	**232**	995.149	12.246	245.163	890.656	995.149	995.149	266.209
	0.0120	(2.30)	330.953	8.282	109.511	299.603	444.519	334.753	108.725
	0.0140		142.850	6.113	53.516	133.565	217.228	149.235	53.918
	0.0100	**263**	438.821	19.821	461.895	410.332	438.821	438.821	460.552
	0.0120	(2.60)	163.465	11.822	166.027	153.315	157.733	163.960	166.305
	0.0140		78.109	7.966	75.198	74.752	71.441	79.942	74.979

Conclusion

In constructing attribute control charts for monitoring zero-inflated processes, an appropriate approach is to fit a ZIP or ZIB model to the zero-inflated count. The parameter value estimated from the model is then used to construct the required c or np charts. However, the resulting chart is still likely to have an ARL_0 value much smaller than the desired value due to the poor coverage probability of its control limit. Our study reveals that in some cases, a simple two-of- two control rule can also be used to enhance the performance of the c - chart.

In the following we introduce one more application of zero-inflated models in process capability index.

6.3 Process Capability Index for Zero-Inflated Poisson Process

Process capability indices have been of interest for the researchers in the recent years. Most of the indices are based on the assumption of normality of process. Indices for non-normal process distributions have also been proposed in the literature. An extensive review of various indices can be found in Kotz and Johnson (2002), Kotz and Lovelace (1998), Spiring et al. (2003) have provided a good review on the bibliography on process capability indices. Borges and Ho (2001) have provided a capability index based on fraction defective. Clements (1989) has given process capability computations for non-normal distributions. Kane (1986), Kotz and Johnson (1993) have studied process capability indices. Pearn and Chen (1995) have proposed the estimating process capability indices for non-normal pearsonian populations. Perakis and Xekalaki (2002) have studied a process capability index based on the proportion of conformance. Yeh and Bhattacharya (1998) have given the robust process capability index. Perakis and Xekalaki (2005) have proposed a new process capability index useful for both the discrete and continuous processes. Further Perakis and Xekalaki (2005) have provided a process capability index for Poisson and attribute data. Their indices are based on maximum likelihood estimate of the Poisson parameter as well as on minimum variance unbiased estimator (MVUE). A simulation study performed by them reveals that indices based on maximum likelihood estimates perform better than the one based on MVUE.

In the recent years, due to adoption of technology, production processes produce extremely good products. Therefore, zero- inflated models have been found useful in modeling production process data. Xie et al. (2001) have proposed control limits based on zero-inflated Poisson model. Naturally, process capability indices (PCI) which are based on Poisson distribution need to be updated, to take an account of zero-inflated behavior property of process.

In the present study, we modify the PCI provided by Perakis and Xekalaki (2005) so as to take an account of inflation at zero in the process. If the process is zero-inflated, then the use of usual Poisson distribution results in underestimating the parameter value. We throw some light on this aspect by providing some numerical study. The proposed index here involves two parameters. Therefore, maximum likelihood estimators for the same have been used. The study of performance of the proposed index has also been taken up.

Suppose X (usually number of defects) denotes the quality characteristic under study which follows Poisson distribution with parameter θ. Let U be the upper tolerance specified by the manufacturer on the number of defects. The upper process capability index (C_{PCU}) defined by Perakis and Xekalaki (2005) is given by

$$C_{PCU} = \frac{1-p_0}{1-p},$$

$$C_{PCU} = \frac{0.0027}{1-p}, \qquad \ldots(6.3.1)$$

where p_0 is the minimum allowable proportion of conformance of the examined process and p is the proportion of conformance which is given by

$$p = p(X < U) = \sum_{x=0}^{U-1} \frac{e^{-\theta}\theta^{x}}{x\,!} \qquad \ldots(6.3.2)$$

It is easy to see that p can be expressed as

$$p = p(\chi^2_{2U} > 2\theta),$$

$$\text{or} \qquad 1 - p = p(\chi^2_{2U} < 2\theta), \qquad \ldots(6.3.3)$$

where χ^2_k denotes the Chi square random variable with k degrees of freedom. Therefore, we have

$$C_{PCU} = \frac{0.0027}{p(\chi^2_{2U} < 2\theta)}. \qquad \ldots(6.3.4)$$

In the recent years, awareness about quality of any production process has been enhanced to great extent and best quality products are being manufactured by adopting technological innovations. Therefore, production processes contain significant number of zero defectives in the production runs. To accommodate this behavior of the process, zero-inflated distributions are being used to model process data. ZIP distribution is one of such distributions.

In the light of presence of inflation in the process distribution, C_{PCU} as defined in Eq. (6.3.1) is modified and it is now given by

$$C^{Z}_{PCU} = \frac{0.0027}{\pi\; p(\chi^2_{2U} < 2\theta)}. \qquad \ldots(6.3.5)$$

We note that, if $\pi = 1$, then C^{Z}_{PCU} coincides with C_{PCU}. Since the index defined in Eq. (6.3.5) contains unknown parameters π and θ. In the following we discuss the estimation of these parameters.

Estimation of C_{PCU}^{Z}

Suppose a random sample $X_1,\ X_2,...,X_n$ is available from the ZIP process. We obtain π and θ in Eq. (6.3.5) by their respective mles.
Substituting $\hat{\pi}$ and $\hat{\theta}$ we get MLEs for $\hat{C}_{PCU}^{Z}$ as

$$\hat{C}_{PCU}^{Z} = \frac{0.0027}{\hat{\pi}\ p(\chi_{2U}^{2} < 2\hat{\theta})}. \qquad \ldots(6.3.6)$$

Assuming asymptotic normality of $(\hat{\pi},\hat{\theta})$, we get $(\hat{\pi},\hat{\theta})' \sim AN_2(\xi',\Sigma)$, where $\xi = (\pi,\theta)$ and $\Sigma = I^{-1}(\pi,\theta)'$ and entries of Σ are given by $\Sigma = (\sigma_{ij})$,

$$\sigma_{11} = \left(\frac{n\left(1-e^{-\theta}\right)}{\pi\left(1-\pi+\pi\,e^{-\theta}\right)}\right)\ ,\ \sigma_{22} = \left(n\pi\left(\frac{(\pi-1)\,e^{-\theta}}{1-\pi+\pi\,e^{-\theta}}+\frac{1}{\theta}\right)\right) \text{ and}$$

$$\sigma_{12} = \sigma_{21} = \left(\frac{n\,e^{-\theta}}{\left(1-\pi+\pi\,e^{-\theta}\right)}\right). \qquad \ldots(6.3.7)$$

To check the performance of proposed process capability index, using ZIP model with the one given by Perakis and Xekalaki (2005), we tabulate the numerical values of the process capability index for U=20 and $\theta = 8,9,...,15$.

Table : 6.3.1

Numerical values of the process capability index for Poisson model and Zero- Inflated Poison model for U=20

θ	C_{PCU}	p	π	C_{PCU}^{Z}
8	10.67449	0.999747	0.4	26.68623
			0.5	21.34899
			0.6	17.79082
			0.7	15.24928
9	2.55693	0.998944	0.4	6.39232
			0.5	5.11386
			0.6	4.26155
			0.7	3.65275
10	0.781625	0.996546	0.4	1.95406
			0.5	1.56325
			0.6	1.30270
			0.7	1.11660
11	0.290652	0.990711	0.4	0.72663
			0.5	0.58130
			0.6	0.48442
			0.7	0.41521
12	0.126881	0.97872	0.4	0.31720
			0.5	0.25376
			0.6	0.25376
			0.7	0.21146
13	0.063278	0.957331	0.4	0.15819
			0.5	0.12655
			0.6	0.10546
			0.7	0.09039
14	0.035292	0.923495	0.4	0.08823
			0.5	0.07058
			0.6	0.05882
			0.7	0.05041
15	0.021638	0.875219	0.4	0.05409
			0.5	0.04327
			0.6	0.03606
			0.7	0.03091

It is clear that if we ignore inflation in the process then the status of the process is underestimated. As the amount of inflation increases (that is π decreases), process capability index shows goodness of the process, which is quite logical.

6.4 Simulation Study

In order to test the performance of the proposed estimator, a simulation study is conducted. In the simulation study 10,000 random samples were generated from the ZIP distribution for various values of the parameters θ, π and for five different sample sizes (25, 50, 100, 200, 400). Two alternative values of U (10 and 20) are chosen so as to detect the influence of all these factors on the behaviors of the estimator. Table 6.4.1 gives the process capability index for U=10 and for different values of inflation parameter π=0.4, 0.5, 0.6, 0.7 and 1.0. Table 6.4.2 gives the process capability index for U=20 and for different values of inflation parameter π=0.4, 0.5, 0.6, 0.7 and 1.0.

In the present study, we propose an estimator for process capability index, when the process distribution is zero-inflated Poisson process. It is observed that ignorance of inflated behavior of the data leads to underestimation of the process capability. It is recommended that, in the presence of zero inflation in the data, index using appropriate zero-inflated distribution gives a better status of the process in terms of the process capability.

Table : 6.4.1

Monte Carlo estimates of C^{Z}_{PCU} (U=10)

θ	C_{PCU}	p	π	Sample size				
				$n=25$	$n=50$	$n=100$	$n=200$	$n=400$
				C^{Z}_{PCU}				
3	2.4490	0.9988	0.4	36.3744	17.6975	9.1900	7.0426	6.6714
			0.5	22.1845	8.5075	6.4212	5.7293	5.2333
			0.6	13.5368	7.4402	4.9704	4.5215	4.3121
			0.7	8.5782	4.9056	4.4660	3.8028	3.6683
			1.0	4.3826	3.2294	2.8441	2.6281	2.5074
4	0.3320	0.9918	0.4	3.6159	1.5052	1.0235	0.9204	0.8719
			0.5	1.9067	0.9231	0.7951	0.7186	0.6898
			0.6	1.0140	0.7105	0.6231	0.5830	0.5697
			0.7	0.7634	0.6089	0.5426	0.4995	0.4850
			1.0	0.4684	0.3862	0.3530	0.34254	0.3409
5	0.0848	0.9681	0.4	0.4180	0.2797	0.2395	0.2261	0.2173
			0.5	0.3192	0.2132	0.1917	0.1774	0.1737
			0.6	0.2058	0.1715	0.1511	0.1455	0.1435
			0.7	0.1708	0.1402	0.1310	0.1251	0.1220
			1.0	0.1025	0.0931	0.0900	0.0859	0.0852
6	0.0321	0.9161	0.4	0.1478	0.0980	0.0892	0.0849	0.0816
			0.5	0.0938	0.0729	0.0695	0.0670	0.0652
			0.6	0.0679	0.0608	0.0571	0.0553	0.0543
			0.7	0.0572	0.0500	0.0473	0.0468	0.0464
			1.0	0.0362	0.0333	0.0332	0.0324	0.0324
7	0.0159	0.8304	0.4	0.0584	0.0468	0.0430	0.0412	0.0405
			0.5	0.0400	0.0350	0.0337	0.0323	0.0320
			0.6	0.0328	0.0286	0.0275	0.0270	0.0267
			0.7	0.0258	0.0244	0.0232	0.0230	0.0229
			1.0	0.0170	0.0167	0.0161	0.0161	0.0160

Table : 6.4.2

Monte Carlo estimates of C_{PCU}^{Z} (U=20)

θ	C_{PCU}	p	π	Sample size				
				$n=25$	$n=50$	$n=100$	$n=200$	$n=400$
				C_{PCU}^{Z}				
8	10.670	0.9997	0.4	202.5163	51.7073	36.1867	30.5956	28.1917
			0.5	88.1043	38.7033	26.9503	24.3865	22.6081
			0.6	49.7162	26.1144	21.5954	19.2383	18.7718
			0.7	31.5318	21.1251	17.9216	16.9967	15.7911
			1.0	17.1357	13.6036	12.0765	11.4464	11.0742
			P&X*	17.3033	13.4535	11.8845	11.2921	10.9865
9	2.557	0.9989	0.4	34.8076	10.8434	8.2225	7.3380	6.7527
			0.5	12.2786	7.1280	6.1069	5.6066	5.2886
			0.6	8.2793	5.9664	5.1004	4.6108	4.4785
			0.7	6.2031	4.9343	4.1165	3.9251	3.7945
			1.0	3.5835	3.1612	2.7624	2.6834	2.6305
			P&X	3.6990	3.0711	2.7960	2.6742	2.6136
10	0.782	0.9965	0.4	5.7521	2.8923	2.4506	2.1705	2.0661
			0.5	3.2385	2.1884	1.8326	1.6894	1.6247
			0.6	2.5451	1.7080	1.4940	1.3903	1.3340
			0.7	2.0179	1.3468	1.2335	1.1875	1.1540
			1.0	1.0530	0.9110	0.8528	0.8115	0.7707
			P&X	1.0494	0.8949	0.8382	0.8078	0.7940
11	0.291	0.9907	0.4	1.8939	1.1086	0.8501	0.7848	0.7558
			0.5	1.0602	0.7531	0.6716	0.6141	0.5963
			0.6	0.8453	0.5908	0.5503	0.5165	0.4987
			0.7	0.6054	0.4864	0.4428	0.4338	0.4234
			1.0	0.3673	0.3303	0.3081	0.3007	0.2929
			P&X	0.3657	0.3240	0.3066	0.2989	0.2942

Table : 6.4.2 continued…

θ	C_{PCU}	p	π	Sample size				
				$n=25$	$n=50$	$n=100$	$n=200$	$n=400$
				C_{PCU}^{Z}				
12	0.127	0.9787	0.4	0.5732	0.4105	0.3625	0.3354	0.3266
			0.5	0.4045	0.3135	0.2761	0.2637	0.2615
			0.6	0.3176	0.2518	0.2309	0.2203	0.2162
			0.7	0.2331	0.2131	0.1927	0.1872	0.1838
			1.0	0.1517	0.1374	01321	0.1294	0.1282
			P&X	0.1517	0.1384	0.1326	0.1296	0.1282
13	0.063	0.9573	0.4	0.2539	0.2072	0.1728	0.1660	0.1615
			0.5	0.1990	0.1496	0.1360	0.1322	0.1282
			0.6	0.1445	0.1225	0.1108	0.1094	0.1066
			0.7	0.1162	0.0991	0.0961	0.0919	0.0908
			1.0	0.0712	0.0672	0.0647	0.0651	0.0638
			P&X	0.0728	0.0677	0.0653	0.0644	0.0638
14	0.035	0.9235	0.4	0.1304	0.1033	0.0956	0.0936	0.0893
			0.5	0.0925	0.0793	0.0753	0.0728	0.0716
			0.6	0.0775	0.0658	0.0627	0.0610	0.0593
			0.7	0.0609	0.0550	0.0527	0.0510	0.0508
			1.0	0.0407	0.0371	0.0358	0.0354	0.0354
			P&X	0.0395	0.0373	0.0363	0.0358	0.0355
15	0.022	0.8752	0.4	0.0947	0.0644	0.05668	0.0564	0.0549
			0.5	0.0556	0.0488	0.0454	0.0444	0.0438
			0.6	0.0438	0.0392	0.0377	0.0365	0.0362
			0.7	0.0356	0.0335	0.0318	0.0314	0.0310
			1.0	0.0238	0.0227	0.0219	0.0219	0.0216
			P&X	0.0236	0.0226	0.0221	0.0219	0.0218

* P&X is the process capability index given by Perakis and Xekalaki (2005)

Example:

The data below is the read write errors discovered in a computer hard disk in a manufacturing process. A set of defect count from a manufacturing process (Xie et al. 2001).

0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	6	0	9
11	0	1	2	0	0	0	0	0	0	0	0	3	3	0	0	5	0	15	6
0	0	0	4	2	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	0	0	0	0	75	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0
0	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
0	0	1	0	0	0	0	0												

In this data set it can be see that, the data set contains many samples with no non-conformities. From the data set we have $n = 208$, $\overline{Y} = \frac{242}{208} = 1.163462$, and the maximum likelihood estimates are $\hat{\pi} = 0.1346$ and $\hat{\theta} = 8.6413$**.** The overall ZIP model for the data set is

$$P(X = x) = \begin{cases} (1\text{-}0.1346)\text{+}0.1346\ e^{-8.4613} & for\ \ x = 0\,, \\ \\ \dfrac{0.1346\, e^{-8.4613}\, 8.4613^{x}}{x!} & for\ \ x = 1,2,3.... \end{cases}$$

The process capability index for the above example using Poisson model and zero-inflated Poisson model are tabulated in the Table 6.5.3.

Table : 6.4.3

U	Poisson Model C_{PCU}	ZIP model C_{PCU}^{Z}
5	0.002898	0.021531
6	0.003137	0.023305
7	0.003559	0.026444
8	0.004269	0.031717
9	0.005441	0.040423
10	0.007388	0.054890

It is recommended that, in the presence of zero inflation in the data, index using appropriate zero-inflated distribution gives a better status of the process in terms of the process capability. Though here we study process capability index C_{PCU}, other indices can also be modified appropriately in order to take an account of inflation in the process.

Future Plans:

1. There is a good scope for extending the results reported here for multivariate set up. Attempts for the same will be made. A problem of interest in the multivariate set will be testing for independence. LRT and Wald's tests will be developed for the same. Confidence intervals for the same will also be studied.
2. Instead of considering models with inflation at a single point, it will be of interest to study models having inflation at more than one point. Inflation may exist even for a subset of the support. We propose to study such models, which will have practical applications.