ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

CONTROL CHARTS
FOR POISSON DATA

By

Evangelia-Lydia {Andreas}. Athanasiou

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece
October 2009

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ ΓΙΑ ΔΕΔΟΜΕΝΑ POISSON

Ευαγγελία-Λυδία Ανδρέα Αθανασίου

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Οκτώβριος 2009

ACKNOWLEDGEMENTS

Many thanks to my supervisor Professor A. Kostaki as well as to Professor S. Psarakis for their help and support.

VITA

Lydia Athanasiou was born in Athens in 1981. After graduating from the German School of Athens, she studied Banking and Finance at the University of Piraeus. She spent one semester as an exchange student at the University of Hamburg. In 2003 she completed her Bachelor’s and started working in the financial department of a logistics company. In 2005 she entered the Athens University of Economics and Business to study Statistics. From 2007 she works as a Credit Risk Analyst.

ABSTRACT

Evangelia-Lydia Athanasiou

Control Charts for Poisson Data

October 2009

This thesis is a review of Control Charts for Data following the Poisson distribution. Firstly, the basic control chart theory and charts for all types of data are being introduced. Then, Shewhart control charts for data following the Poisson distribution are being presented: The modified u Control Chart as well as the c Chart for Nonconformities which is being compared to the Poisson Moving Average chart. The Zero Inflated Poisson Model which monitors processes with excessive 0 counts and the generalized zero inflated Poisson distribution, an extension of the ZIP model are introduced. Furthermore, several EWMA as well as CUSUM control charts (the Double EWMA, the PGWMA Chart, and finally FIR Poisson EWMA control charts) are examined. As far as CUSUM charts are concerned, the conditional and marginal performance of a Poisson CUSUM chart when the parameter is unknown and the CUSUM control chart based on the Poisson distribution compounded by a Geometric distribution are examined. Finally, Control Charts for Multivariate Poisson distributions are presented.

.

ΠΕΡΙΛΗΨΗ

Ευαγγελία-Λυδία Αθανασίου

Διαγράμματα Ελέγχου για Δεδομένα Poisson

Οκτώβριος 2009

Η παρούσα εργασία αποτελεί μια επισκόπηση Διαγράμματων Ελέγχου που αφορούν δεδομένα τα οποία ακολουθούν την Poisson κατανομή. Αρχικά αναφέρονται και αναλύονται τα βασικά διαγράμματα ελέγχου και η θεωρία αυτών. Έπειτα, εξετάζεται το τροποποιημένο διάγραμμα u και παρουσιάζεται μια σύγκριση του διαγράμματος c με το διάγραμμα του κινητού μέσου (moving average). Επίσης, γίνεται αναφορά σε ένα απλό και ένα γενικευμένο zero inflated Poisson διάγραμμα. Ακόμη, παρουσιάζονται διάφορα EWMA και CUSUM διαγράμματα ελέγχου: Διπλό (Double) EWMA, PGWMA και FIR PEWMA καθώς επίσης εξετάζονται η οριακή και δεσμευμένη απόδοση του Poisson CUSUM και ένα διάγραμμα CUSUM για μια σύνθετη Poisson κατανομή. Τέλος, γίνεται αναφορά σε διαγράμματα ελέγχου για την πολυμεταβλητή Poisson κατανομή.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1.

INTRODUCTION

Quality can be defined in many ways. Many people conceptualize quality as one or more quality characteristics a product should possess. However, as one of the most important consumer decision factors, quality can be more precisely defined. Quality has actually many dimensions. Garvin provides 8 components of quality:

1. Performance: Does the product perform specific actions for the intended job?
2. Reliability: How often does the product fail and requires service/repair?
3. Durability: Does the product last for a satisfactorily long period?
4. Serviceability: If the product requires service, how quickly and economically can it be repaired.
5. Aesthetics: Style, shape, color and the visual appeal of the product in general are often an important factor that is taken under consideration.
6. Features: Does the product do things that its competitors do not?
7. Perceived Quality: What is the past reputation of the product and the company that produces it?
8. Conformance to Standards: Is the product made exactly as intended by the designers

Improving quality is one of the main targets in modern businesses and statistical methods play a central role in quality improvement efforts. The most important reason that makes it difficult for businesses to provide the consumer products that are always identical from unit to unit is the variability that exists in every product.

In order to achieve stability and minimize the variability of a process a selection of tools called statistical process control (SPC) is used. The seven major tool of SPC are:

1. Histogram
2. Check Sheet
3. Pareto Chart
4. Cause and effect diagram
5. Defect concentration diagram
6. Scatter Diagram
7. Control Chart

The most technically sophisticated tool is the control chart developed in the 1920s by Dr. Walter A. Shewhart of the Bell Telephone Laboratories.

In every process an amount of natural variability will always exist. If a process is operating with only chance causes of variation the process is said to be ‘in statistical control’. If other causes of variability that are not part of the chance exist the process that is operating in the presence of assignable causes is ‘out of control’.

The objective of SPC and in specific control charts, is to detect the occurrence of assignable causes of process shifts as quickly as possible, so as for corrective actions to take place before many non-conforming items are produced. Control charts can also be used to estimate parameters of a production process and to determine process capability. They can also provide useful information so as to improve the process.

This thesis presents Control Charts for Data following the Poisson distribution. The layout is as follows: The second Chapter presents the basic control chart theory and charts for all types of data. More specific all Shewhart control charts as well as the CUSUM and EWMA control charts are being presented.

Chapter 3 presents Shewhart control charts for data following the Poisson Distribution. More specific the C Chart for Nonconformities is being compared to the Poisson Moving Average chart and the modified U Control Chart is also being presented. Finally, the Zero Inflated Poisson Model is introduced in order to monitor processes with excessive 0 counts. A model was developed as a zero defect process subject to random shocks. The random shock occurs with probability p and upon the occurrence of a random shock, non-conformities can be found. This number of nonconformities follows the Poisson distribution. A generalized zero inflated Poisson distribution, an extension of the ZIP model is also presented. The GZIP model is also a particular form of the Poisson distribution.

In the fourth Chapter several EWMA as well as CUSUM control charts are being examined. Not only the simple EWMA, but also the Double EWMA chart is being presented. Moreover, the PGWMA Chart, a generalized charting model of which the PEWMA chart and c-chart are special cases is included in this chapter as well as other PGWMA control charts. The conditional and marginal performance of a Poisson CUSUM chart when the parameter is unknown is also examined in this Chapter. Finally the CUSUM control chart based on the Poisson distribution compounded by a Geometric distribution is presented.

Chapter 5 presents a model-based control chart for a multivariate Poisson with measurable inputs. A multivariate Poisson chart in order to monitor the correlated multivariate Poisson count data is also presented.

CHAPTER 2.

CONTROL CHARTS FOR VARIABLES

2.1 Introduction

Quality characteristics can be expressed in terms of a numerical measurement. A single measurement quality characteristic, such as a dimension, weight or volume is called a variable. It is usually necessary to monitor both the mean value of the quality characteristics and its variability. Control of the process average or mean quality level is usually done with the control chart for means or the x-chart. Process variability can be monitored with either a control chart for the standard deviation, called the S chart or the chart of the range called an R chart.

2.2 x and R charts

Suppose that a quality characteristic is normally distributed with mean μ and standard deviation σ, where both μ and σ known. If x_1, $x_2, \dots, x_n$ is a sample of size n, then the average of this sample is:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}, \qquad \text{with } \bar{x} \sim N\,(\mu,\ \sigma/\sqrt{n})$$

Any sample mean with probability 1-a will fall between:

$\mu + z_{\alpha/2}\,\sigma/\sqrt{n}$ and $\mu - z_{\alpha/2}\,\sigma/\sqrt{n}$

It is customary to replace $z_{\alpha/2}$ with 3, so we have the so called 3-sigma control limits. This implies that the type I error probability is $\alpha=0.0027$.

If a sample falls outside of these limits, this is an indication that the mean of the process is no longer μ.

These results are approximately correct even if the distribution of the process is not normal, due to the central limit theorem.

Unfortunately μ and σ aren't usually known so we have to estimate them from samples which we know that are in control. These estimates should be based on 20-25 samples. Suppose the number of the samples is m and the number of observations n (usually small). Then the mean of each sample is being calculated ($\bar{x}_1, \bar{x}_2, \ldots \bar{x}_m$). The best estimator of μ, the process average, is:

$\bar{\bar{x}} = \dfrac{\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_m}{m}$, $\bar{\bar{x}}$ will be the center line of the $\bar{x}$ chart.

Then we need an estimator of σ, which we can estimate from the ranges of the m samples. The range of a sample is the difference between the largest and the smallest observations: $R = x_{max} - x_{min}$

We get m ranges. The average range is: $\bar{R} = \dfrac{R_1 + R_2 + \ldots + R_m}{m}$

So the control limits for the x chart are:

$$UCL = \bar{\bar{x}} + A_2 \bar{R}$$

$$Center\ line = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - A_2 \bar{R}$$

A_2 is a constant that depends only on the size of the sample.

If we plot the values of the sample range R on a control chart we would have the below control limits for the R chart:

$$UCL = D_4 \bar{R}$$

$$Center\ line = \bar{R}$$

$$LCL = D_3 \bar{R}$$

D_3 and D_4 are constants that depend only on the size of the sample.

2.3 Development of equations for computing the control limits

There is a relationship between the range and the standard deviation of sample from a normal distribution. The random variable W=R/s is called the relative range. The parameters of the distribution of W are a function of the sample size n. The mean of W is d_2. Therefore an unbiased estimator of σ is $\hat{\sigma} = \frac{\overline{R}}{d_2}$

We use this unbiased estimator to calculate the control limits of the $\overline{x}$ chart.

Using the same relationship W=R/s an unbiased estimator of the standard deviation of R can be obtained. We have R=Ws. The standard deviation of W, d_3 is a function of n. The standard deviation of R is $\sigma_R = d_3\sigma$. We estimate σ_R by

$$\hat{\sigma}_R = d_3 \frac{\overline{R}}{d_2}.$$

2.4 Phase 1 and Phase 2 analysis

During the phase 1 analysis (retrospective study of past data) we use the so called *trial control limits*, which allow us to check whether the process was in control when m initial samples were selected. To test this hypothesis we plot the values of $\overline{x}$ and R from each sample on the control charts. If all the points are inside the control limits, then we could say that the process was in control in the past and the control limits that we have obtained can be used for controlling future production.

If one or more points fall outside these control limits it would be useful to revise these limits. If we can find a cause for the outlier, we can remove it and recalculate the control limits. We can remove the outlier even if there isn't a specific cause for it, because the point(s) outside the trial control limits are likely to have been drawn from a probability distribution characteristic of an out-of-control state. Or we could retain the point(s) considering the control limits as appropriate.

If the point(s) represent an out of control condition the control limits would be too wide. However, if there are only one or two such points, these will not distort the control chart significantly. The initial set of control limits should be revised periodically.

Once a set of reliable control limits is established, we use the control chart for monitoring future production. This is the so called phase 2 of control chart usage.

2.5 Specification Limits and Control Limits

There is no relationship between the Control Limits and the so called Specification Limits. The Control Limits are also called Natural Tolerance Limits of the process and are as described above +3σ or -3σ the mean of the process. The specification limits are determined externally (from engineers, the management, the customer) and they don't have anything to do with the Control Limits of the process.

2.6 Between- and within sample variability

The x chart measures the average quality level in the process; the R chart measures the variability within the samples. The $\bar{x}$ chart monitors between sample variability and the R chart measures within sample variability.

At this point we could point out, that when we calculate an estimator of the standard deviation σ to determine the control limits; we are interested in the standard deviation within each sample. So, it would be false to use the quadratic estimation:

$$S=\sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{\overline{x}})^2}{mn-1}}\ ,$$

because if sample means differ, then S will be to large. This S would combine both between and within data variability, whereas the control limits should be based only on within data variability.

2.7 The problem of control chart design

A unique solution for all the problems of control charts design does not exist. To specify the size of the samples the frequency of sampling and the control limit width one should be aware of the cost of sampling, the costs of investigating and correcting the process, when it is out of control and the cost of an out-of-control product. An economic decision model should be designed to solve this problem.

For large process shifts (larger than 2σ) small sample sizes of n=4, 5 or 6 are being recommended, for smaller process we could take bigger sample sizes.

As far as the R chart is being concerned, it is not very sensitive to shifts in the process standard deviation for small sample sizes. Unfortunately if we take larger samples, this would result less efficient estimators of the standard deviation.

The operating characteristic curves of the $\overline{x}$ and R charts can help us in choosing the right sample size. Another factor that influences the choice of the sample size and the sample frequency is the rate of production.

2.8 Interpretation of control charts

A control chart can include an out of control condition even though all points are in control. Sometimes the points of the chart show a systematic behavior, a pattern, which we found usually during the phase 1 analysis. Their elimination is crucial in bringing process into control. Such patterns are a cyclic pattern, a mixture (when points tend to fall outside control limits and only few points are near the center line), a shift in process level or a trend.

2.9 Nonnormality of control charts.

A fundamental assumption in the development of the above control charts is that the distribution of the quality characteristic is normal. Sometimes this assumption is not valid. If we know which distribution is appropriate for our data, it is possible to derive the sample distributions and the exact probability limits for the control charts. This is usually very difficult. Sometimes the form of the distribution of the data is often unknown to us. For these reasons we use the normal theory results. The normality assumption can be employed unless the data are extremely nonnormal.

The distribution of R is not symmetric, even when sampling from the normal distribution. So, symmetric 3σ limits are only an approximation, and the α risk is not 0.0027.

2.10 The operating characteristic function and the Average Run Length (ARL)

The ability of the $\bar{x}$ and R charts to detect shifts in the process is described from the operating characteristic curves. To understand the significance of the OC curves, we should explain the meaning of the β risk. The β risk is the probability of not detecting a shift in our process when we take a sample. For example if the mean shifts from μ_0 to another value $\mu_1=\mu_0+2\sigma$, the β risk is:

$$\beta=P(LCL\leq \bar{x}\leq UCL \mid \mu=\mu_1=\mu_0+2\sigma)$$

Since $\bar{x}\sim N(\mu,\sigma^2/n)$ and $UCL=\mu+3\sigma/\sqrt{n}$, $LCL=\mu-3\sigma/\sqrt{n}$ the β risk is:

$$\beta=\Phi\left[\frac{UCL-(\mu_0+2\sigma)}{\sigma/\sqrt{n}}\right]-\Phi\left[\frac{LCL-(\mu_0+2\sigma)}{\sigma/\sqrt{n}}\right]=\Phi(3-2\sqrt{n})-\Phi(-3-2\sqrt{n})$$

The probability that shift will be detected on the 1st sample is 1-β.

To construct the OC curve for the $\bar{x}$ chart, we plot the β-risk on the vertical axis and the magnitude of the shift in standard deviation units on the horizontal axis for various sample sizes n.

The probability that the shift will be detected on the 2nd sample is: β(1-β) and on the rth sample is the probability of not detecting the shift on each of the r-1 samples times the probability that the shift will be detected on the rth sample:β^{r-1}(1-β).

The expected number of samples taken before the shift is detected is the Average Run Length: $ARL_1=\sum_{r=1}^{\infty} r\beta^{r-1}(1-\beta) = \frac{1}{1-\beta}$

In general we can say that:

$$ARL=\frac{1}{P(one\ \ point\ \ plots\ \ out\ \ of\ \ control)}$$

$ARL_0=1/\alpha$ for the in-control ARL

$ARL_1=\frac{1}{1-\beta}$ for the out of control ARL

ARL_1 shows the number of samples needed to detect a shift in the process, in other words at which sample we find out that the process is out of control, when the process is out of control. ARL_0 on the other hand, shows the number of samples needed to find out that the process is out of control, when the process is in control, in other words the number of sample needed to have a false alarm. The ARL can be expressed for any Shewhart control chart.

2.11 x and S charts

Sometimes we need to estimate the standard deviation directly and not through the range R. This leads to control charts of for $\bar{x}$ and S, where S is the sample standard deviation. $\bar{x}$ and S charts are preferable for larger sample sizes or for variable sample sizes.

The sequence of steps is the same as in the $\bar{x}$ and R charts, but at this case we should calculate the sample average and the sample standard deviation for each of the samples. We know that an unbiased estimator of the unknown variance σ² of our distribution is $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$.

However, S is not an unbiased estimator of the standard deviation σ. If the underlying distribution is normal, S actually estimates $c_4\sigma$, where c_4 is a constant that depends only on the sample size n. The standard deviation of S is $\sigma\sqrt{1-c_4^2}$.

The center line for the S chart is $c_4\sigma$ and the control limits are:

$$UCL = B_6\sigma$$
$$Center\ Line = c_4\sigma$$
$$LCL = B_5\sigma$$

Where $B_5 = c_4 - 3\sqrt{1-c_4^2}$ and $B_6 = c_4 + 3\sqrt{1-c_4^2}$ depend only on the size of the sample.

If no standard is given for σ, then we must estimate it from past data: If we have taken m samples of size n then and S_i is the standard deviation of the i th sample, the average of all the m standard deviations is $\bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_i$.

$\bar{S}/c_4$ is an unbiased estimator of σ. The parameters of the S chart should be:

UCL $= B_4\bar{S}$

Center line= $\bar{S}$

LCL $= B_3\bar{S}$

Where $B_3=1-\frac{3}{c_4}\sqrt{1-c_4^2}$ and $B_4=1+\frac{3}{c_4}\sqrt{1-c_4^2}$ depend only on the size of the sample.

The control limits for the $\bar{x}$ chart are:

UCL= $\bar{\bar{x}}+A_3\bar{S}$

Center Line= $\bar{\bar{x}}$

LCL= $\bar{\bar{x}}-A_3\bar{S}$,

Where $A_3=3/(c_4\sqrt{n})$ depends only on the size of the sample.

<u>2.12 The x bar and S Control Charts for variable sample size</u>

If the sample sizes are variable we can calculate weighted $\bar{\bar{x}}$ and $\bar{S}$.

$$\bar{x}=\frac{\sum_{i=1}^{m} n_i\bar{x}_i}{\sum_{i=1}^{m} n_i} \quad , \quad \bar{S}=\left[\frac{\sum_{i=1}^{m}(n_i-1)S_i^2}{\sum_{i=1}^{m} n_i-m}\right]^{1/2}$$

We will use the same equations to calculate the Control Limits as for a fixed sample size, but the constants A_3, B_3 and B_4 depend on the sample size of the individual group.

<u>2.13 The Shewhart Control Charts for individual measurements</u>

Sometimes the sample size is n=1. On those cases we use the moving ranges of the observations to estimate the process variability, that is:

$MR_i = |x_i - x_{i-1}|$. We can also illustrate a control chart on the moving range:

$$\text{UCL} = \bar{x} + 3\frac{\overline{MR}}{d_2}$$

$$\text{Center line} = \bar{x}$$

$$\text{LCL} = \bar{x} - 3\frac{\overline{MR}}{d_2}$$

Moving ranges are correlated and this can result a pattern of cycles on the chart. The moving range chart cannot really provide useful information about a shift in process variability. Shifts in the mean also show up in the moving range chart.

2.14 Control Charts for Attributes

The terminology *conforming* or *nonconforming* is being used to give two classifications to some products, either because many characteristics cannot be represented numerically or because we are interested only in the adequacy of a product. These quality characteristics are being called *attributes*. The attributes control charts are the *p chart* or *control chart for fraction non-conforming*, the *control chart for non-conformities* or the *c chart*, which represents the number of defects per unit and the *control chart for non-conformities per unit* or the *u chart*, which represents the average number of nonconformities per unit.

2.15 The control chart for fraction non-conforming

The fraction non-conforming is defined as the ratio of the number of non-conforming items in a population to the total number of items in that population. If an item does not conform to specific standards given for its characteristic, we classify it as non-conforming. We express the faction nonconforming as decimal or as percent non-conforming.

The control chart for fraction nonconforming is based on the binomial distribution. Suppose that the probability, that unit will not conform to specifications is p and that successive units produced are independent. In that case, each unit is the realization is of a Bernoulli random variable with parameter p. If the sample of n units is being selected and D is number of defective-nonconforming units, then D has a binomial distribution with parameters n and p:

$$P(D=x)=\binom{n}{x}p^x(1-p)^{n-x},\ x=0,1,\ldots,n$$

The sample fraction nonconforming is defined as:

$$\hat{p}=\frac{D}{n}$$

The distribution of the random variable $\hat{p}$ can be obtained from the binomial. The mean and the variance of $\hat{p}$ are μ=p and $\sigma_{\hat{p}}^2=\frac{p(1-p)}{n}$ respectively.

T he 3 sigma control limits are for a known fraction nonconforming p or a standard value:

UCL= $p+3\sqrt{\frac{p(1-p)}{n}}$

Center line=p

LCL= $p-3\sqrt{\frac{p(1-p)}{n}}$

In this case we take the samples of n units and compute the fraction nonconforming $\hat{p}$, which we plot on the control chart. If $\hat{p}$ remains within the limits, our process is in control. If a point plots outside the control limits or we can observe a nonrandom pattern in the points of the chart, perhaps there is a shift to a new level.

If the fraction nonconforming isn't known we have to estimate it. We take m samples of size n and count the number of nonconforming units D. If there are D_i nonconforming units in sample i, we compute:

$\hat{p}_i = \dfrac{D_i}{n}$ i=1,…,m the fraction nonconforming for each sample i

$\overline{p} = \dfrac{\sum_{i=1}^{m} D_i}{mn} = \dfrac{\sum_{i=1}^{m} \hat{p}_i}{m}$ the average of these individuals fraction nonconforming

$\overline{p}$ estimates the unknown fraction nonconforming p. These are the 3 sigma control limits, which can be regarded as trial control limits:

UCL= $\overline{p} + 3\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$

Center line= $\overline{p}$

LCL= $\overline{p} - 3\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$

Since the above limits are trial control limits we plot all the sample values p_i to see whether our procedure was in control when we collected the data. If some points are out of control, then we should look for the causes for that and if we find assignable causes, we should discard these points and recalculate the control limits.

If the control chart is based on a standard value trial control limits are unnecessary. However we must be very careful if the desired target value suits the true value p, because if that is not the case our process would seem to be out of control for that p, although it is in control for the true value p.

2.16 The np Control Chart

It is possible to base a control chart on the number nonconforming. This is the so called the np control chart. The parameters of this chart are:

UCL= $np + 3\sqrt{np(1-p)}$

Center line=np

LCL= $np - 3\sqrt{np(1-p)}$

If a standard value for p is unavailable we estimate as we did for the p chart.

2.17 Control Charts for Nonconformities

A nonconforming item contains at least one nonconformity. However a product can have one or more nonconformities and still be conforming. Sometimes we are interested in the number of defects in a product and not in characterising it as conforming or nonconforming. We can develop control charts for either the total number of nonconformities in an item or the average number of nonconformities per item. It is assumed that the occurrence of nonconformities in samples of constant size is well modelled by the Poisson distribution, which means that the probability of occurrence of nonconformities should be a constant and that the number of opportunities for the occurrence of nonconformities has to be infinity. This is not always the case, but the Poisson model works reasonably well.

2.18 Constant sample size

Usually one single product is being inspected of nonconformities, but more products can be inspected also. The nonconformities occur as mentioned above according to the Poisson distribution: $p(x) = \frac{e^{-c} c^{x}}{x!}$ x=0,1,2…

The mean and the variance of the Poisson distribution is the parameter c. If a standard value of c is available then a control chart of 3 sigma limits is:

UCL= $c + 3\sqrt{c}$

Center line= c

LCL= $c - 3\sqrt{c}$

If no standard is given we have to estimate c from a preliminary sample. The estimator of c, $\bar{c}$ is the number of nonconformities in a preliminary sample. The control chart has parameters defined as follows:

UCL= $\bar{c} + 3\sqrt{\bar{c}}$

Center line= $\bar{c}$

LCL= $\bar{c} - 3\sqrt{\bar{c}}$

The above limits should be regarded as trial control limits.

Poisson Data with parameter λ=3 are generated and the corresponding in line c-control chart is presented:

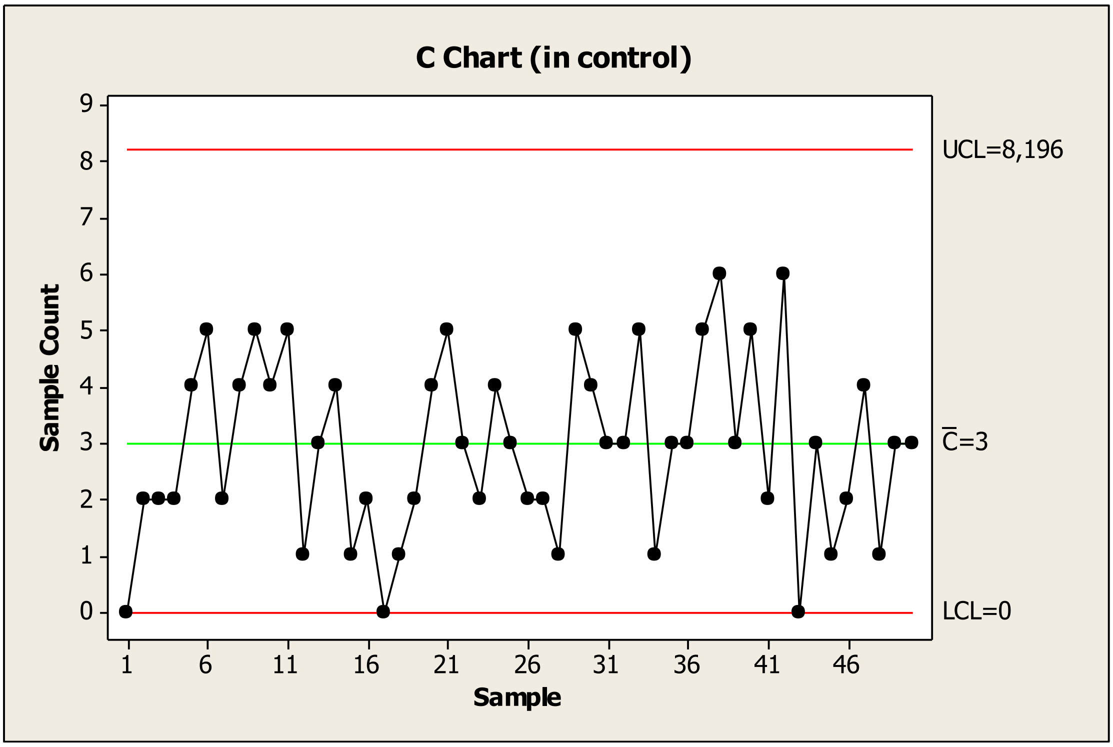

Figure 2.1. In control c chart for simulated Poisson Data with λ=3.

After a shift in the process and (λ=4) c chart detectσ the change of the process slightly after the change.

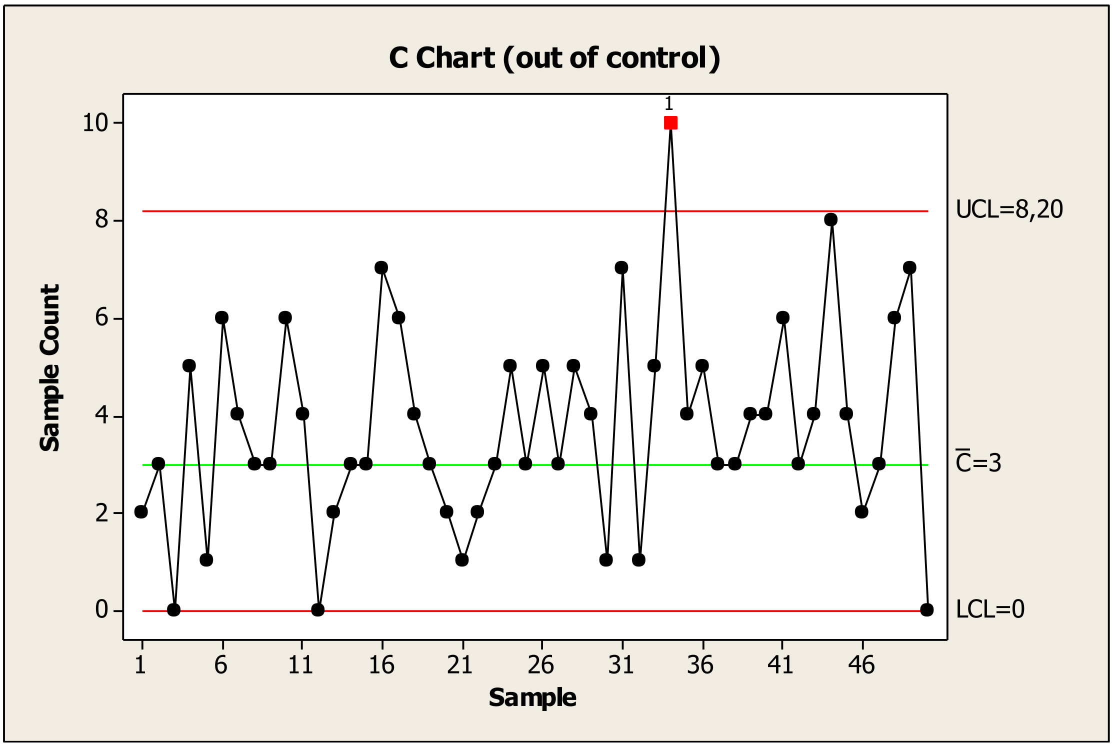

Figure 2.2. Out of control c chart for simulated Poisson Data with λ=4. The process shift is detected.

2.19 The u chart-several inspection units

It is better to use several inspection units in the sample, increasing the area of opportunities for nonconformities. We should choose a sample size large enough to ensure a positive lower control limit and take economic factors into account as well. The size of the sample n does not have to be an integer. If x is the number of nonconformities per inspection units in sample of n units, the average number of nonconformities per inspection unit is:

$u = \frac{x}{n}$, where x is a Poisson random variable

We calculate the average number of defects per unit in a preliminary set of data for m samples: $\bar{u} = \frac{\sum_{i=1}^{m} u_i}{m}$

The control limits for the $\bar{u}$ chart would be:

UCL= $\bar{u} + 3\sqrt{\frac{\bar{u}}{n}}$

Center line= $\bar{u}$

LCL= $\bar{u} - 3\sqrt{\frac{\bar{u}}{n}}$

The above control limits are trial control limits and should be adopted if the points plot within the control limits.

2.20 Variable sample size

Sometimes the number of units in a sample is not a constant, but a variable. If we use a control chart for nonconformities (c chart), the control limits and the center line will vary with the sample size and this would be often very difficult to interpret. It would be better to use a u control chart (nonconformities per unit). This would have one constant center line but the control limits would vary with the sample size. We calculate the average number of defects to the total number of inspection units: $\bar{u} = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} n_i}$

The control limits would be:

UCL= $\bar{u} + 3\sqrt{\frac{\bar{u}}{n_i}}$

Center line= $\bar{u}$

LCL= $\bar{u} - 3\sqrt{\frac{\bar{u}}{n_i}}$

When the sample size is variable we should always use the u chart. Some other possible approaches would be to use control limits based on an average sample size $\bar{n} = \sum_{i=1}^{m} \frac{n_i}{m}$ or use a standardized control chart, plotting the standardized statistic:

$Z_i = \frac{u_i - \bar{u}}{\sqrt{\frac{\bar{u}}{n_i}}}$ on a control chart with

LCL=-3

Center Line=0

UCL= 3

2.21 Cumulative Sum and Exponentially Weighted Moving Average Control Chart

All the charts that had been represented above are called Shewhart Control Charts. The most important disadvantage of the Shewhart Control Chart is that it uses only the information we obtain for the last plotted unit, it ignores the information given about the

sequence of all the point. For that reason Shewhart control charts are insensitive to smaller shifts (<1.5σ).
To detect smaller shifts it is preferable to use the cumulative sum or CUSUM control chart and the EWMA (Exponentially Weighted Moving Average) control chart.

2.22 The CUSUM Control Chart

As we have mentioned the Shewhart Control Chart is very effective for shifts >1.5σ. The cusum chart is a very good alternative if we want to detect small shifts to our process. The cusum control chart uses up all the information we get in the sequence of the points, because it plots the cumulative sums of the deviations of the sample values from a target value. For example, if $\bar{x}_j$ is the average of the j th and μ_0 is the target value for the process mean, we plot the quantity

$$C_i = \sum_{j=1}^{i} (\bar{x}_j - \mu_0)$$

against the sample I and get the cusum control chart. C_i is called the cumulative sum up to and including the i th. As it has been mentioned cusum control charts are more effective, when we are interested in detecting small shifts. In addition, they are very effective for sample size n=1.

If the process remains in control for the target value μ_0, the cusum is random walk with mean zero. If the mean shifts upward or downward the cusum will develop a positive or a negative drill respectively. So, if we can see a trend in the plotted points this should be evidence that the process mean has shifted and we should search for an assignable cause.

Until now we do not have a control chart yet, because although the cusum detects a shift in the process mean we do not have control limits. There are two ways to represent cusums, the tabular (or algorithmic) cusum and the V-mask cusum.

2.23 The Tabular CUSUM

We can construct CUSUMs either for individual observations or for the averages of samples with n>1. First we will construct a CUSUM for individual observations. Let x_i be the i^{th} observation of the process with a distribution with mean μ_0 and standard deviation σ when the process is in control. The mean μ_0 is also called the target value of x. We compute the derivations that are above target with one statistic C^+ , the so called one sided upper CUSUM, and the derivations that are below the target value with another statistic C^-, the so called one sided lower CUSUM. The upper and lower CUSUM are computed as follows:

$$C_i^+ = \max[0, x_i - (\mu_0 + K) + C_{i-1}^+]$$
$$C_i^- = \max[0, (\mu_0 - K) - x_i + C_{i-1}^-]$$

Where the starting values are $C_0^+ = C_0^- = 0$

K is usually called the reference value and it is often chosen halfway between μ_0 and the out-of-control value μ_1 that we are interested in detecting quickly. If $\mu_1=\mu_0+\delta\varsigma$ or $\delta=|\mu_1-\mu_0|/\sigma$, then K=$\frac{\delta}{2}\sigma = \frac{|\mu_1 - \mu_0|}{2}$. C_i^+ and C_i^- accumulate deviations from the target value μ_0 that are greater than K. Both quantities are being reset to zero when they become negative. If either C_i^+ or C_i^-exceed a decision interval H, the process is considered to be out of control.

We can plot C_i^+ and C_i^- versus the sample number. We also plot the decision interval on the chart. A reasonable value for H is 5σ.

Like in Shewhart control charts if a point plots out of control, we search or an assignable cause for that, take actions if necessary and reinitialize the CUSUM at zero. The CUSUM helps us to determine when exactly the shift had occurred. We just have to count backward from the out-of-control signal to the time period when the CUSUM lifted above zero to find the first period following the process shift.

The CUSUM we have described is the so called two-sided CUSUM because it is constructed by C_i^+ and C_i^- two procedures. Sometimes we are interested in an increase or a decrease of a characteristic. In that case it we use the one-sided CUSUM.

The graph below represents an in-control process with λ=3 for simulated data.

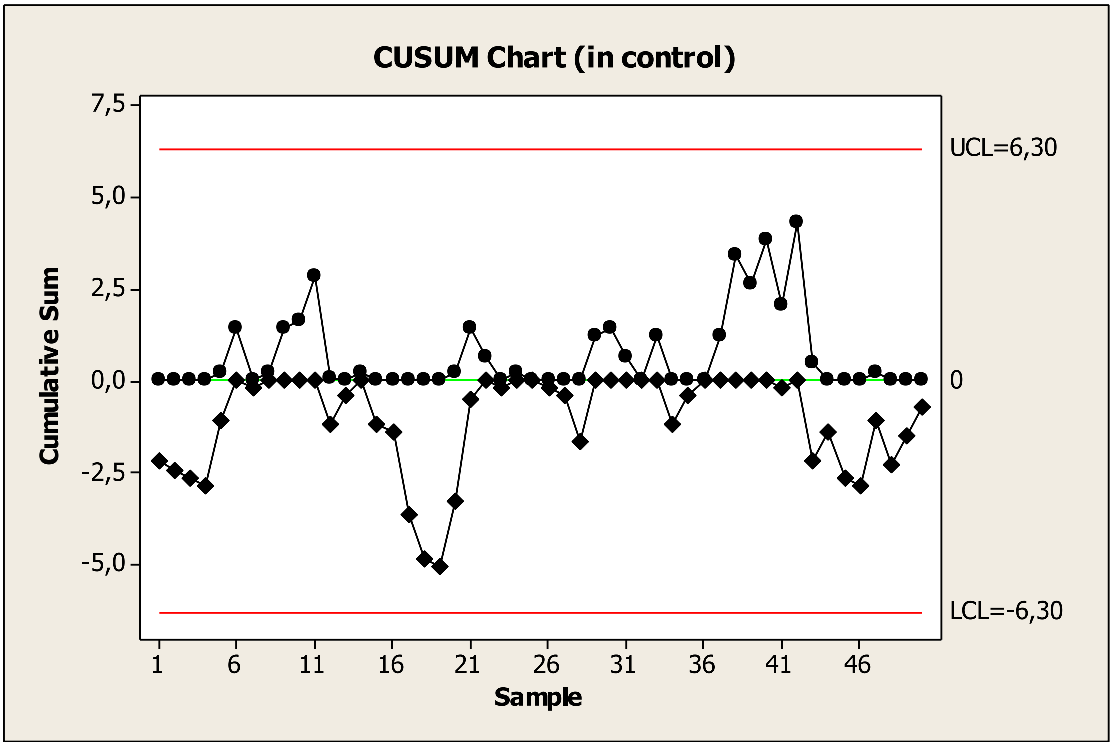

Figure 2.3. In control CUSUM chart for simulated Poisson Data with λ=3.

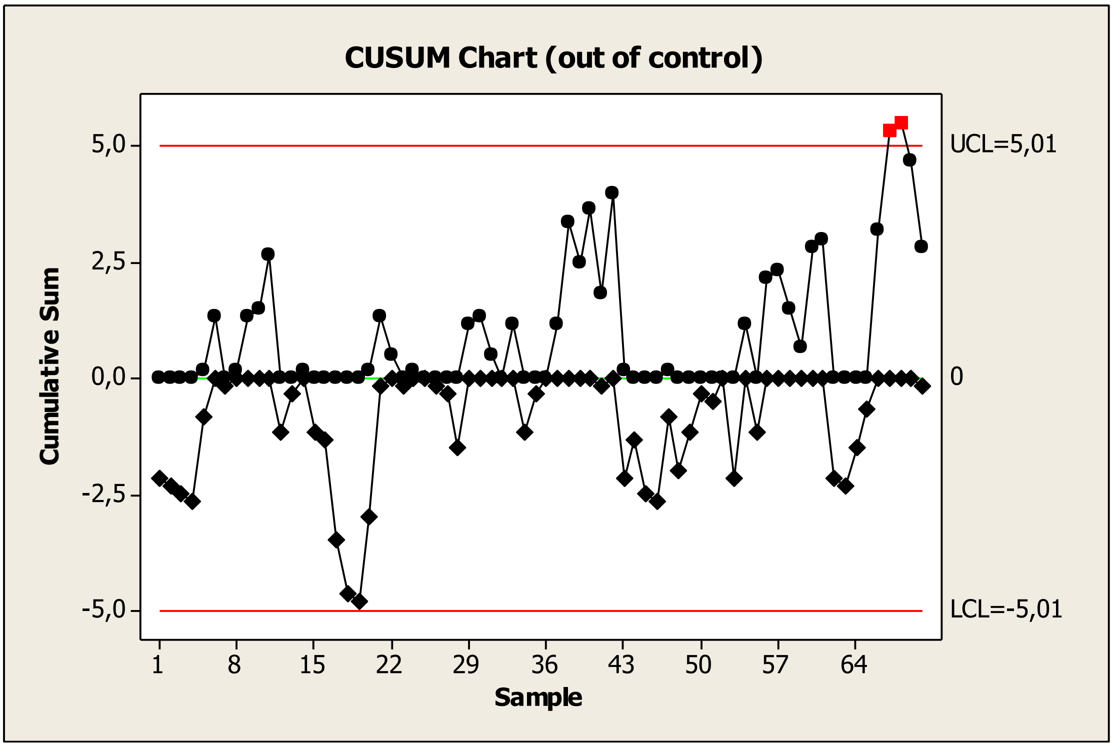

Figure 2.4. Out of control CUSUM chart for simulated Poisson Data with λ=3 that changes to λ=4. The CUSUM detects the change.

2.24 Exponentially Weighted Moving Average Control Chart

The exponentially weighted moving average or EWMA control chart is also suitable for detecting small shifts. It has also the advantage that it is easier to set up in comparison to the cusum. Like the cusum it is used mostly for individual observations. The exponentially weighted moving average is defined us:

$$z_i = \lambda x_i + (1-\lambda) z_{i-1},$$

Where $0<\lambda\leq 1$ is a constant and the starting value for i=1 is the process target $z_0=\mu_0$.

Sometimes the average of the given data is the starting value of the EWMA that is $z_0=\bar{x}$. The EWMA z_i is a weighted average of all previous sample means:

$$z_i=\lambda x_i+ (1-\lambda)\ [\lambda x_{i-1}+ (1-\lambda)\ z_{i-2}]$$

$$= \lambda x_i+ (1-\lambda)\ \lambda x_{i-1}+ (1-\lambda)^2\ z_{i-2}$$

$$z_i = \lambda \sum_{j=0}^{i-1} [(1-\lambda)^j x_{i-j}] + (1-\lambda)^i z_0$$

The weights $\lambda\ (1-\lambda)^j$ decrease geometrically with the age of the sample mean. The weights sum to unity, since $\lambda \sum_{j=0}^{i-1} (1-\lambda)^j = \left[\frac{1-(1-\lambda)^i}{1-(1-\lambda)}\right] = 1-(1-\lambda)^i$

The EWMA can be viewed as weighted average of all past and current observations and it is very insensitive to the normality assumption. If the observations x_i are independent random variables with variance σ^2, the variance of z_i is:

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]$$

The EWMA control chart would be constructed by plotting z_i versus the sample number i. The control limits and the center line are as follows:

UCL= $\mu_0 + 3\sigma\sqrt{\frac{\lambda}{2-\lambda}[1(1-\lambda)^{2i}]}$

CL= μ_0

LCL= $\mu_0 - 3\sigma\sqrt{\frac{\lambda}{2-\lambda}[1(1-\lambda)^{2i}]}$

2.25 The EWMA for Poisson Data

The EWMA can be also used for Poisson data. The exponentially weighted moving average remains unchanged: $z_i = \lambda x_i + (1-\lambda) z_{i-1}$ with $z_0 = \mu_0$, but the control chart parameters are as follows:

UCL $= \mu_0 + A_u\sqrt{\frac{\lambda\mu_0}{2-\lambda}[1-(1-\lambda)^{2i}]}$

Center line=μ_0

LCL $= \mu_0 + A_l\sqrt{\frac{\lambda\mu_0}{2-\lambda}[1-(1-\lambda)^{2i}]}$

A_u and A_l are the upper and lower control limit factors. We use often $A_l = A_u = A$.

The graph below represents an in-control process with λ=3 for simulated data

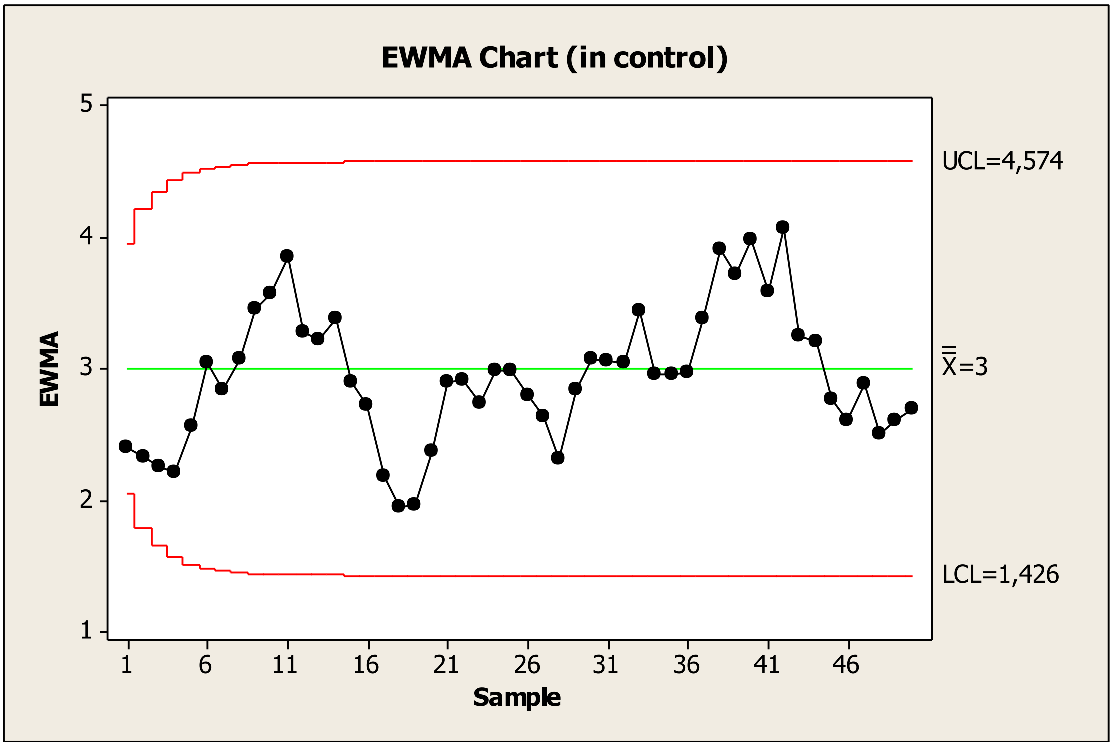

Figure 2.5. In control EWMA chart for simulated Poisson Data with λ=3.

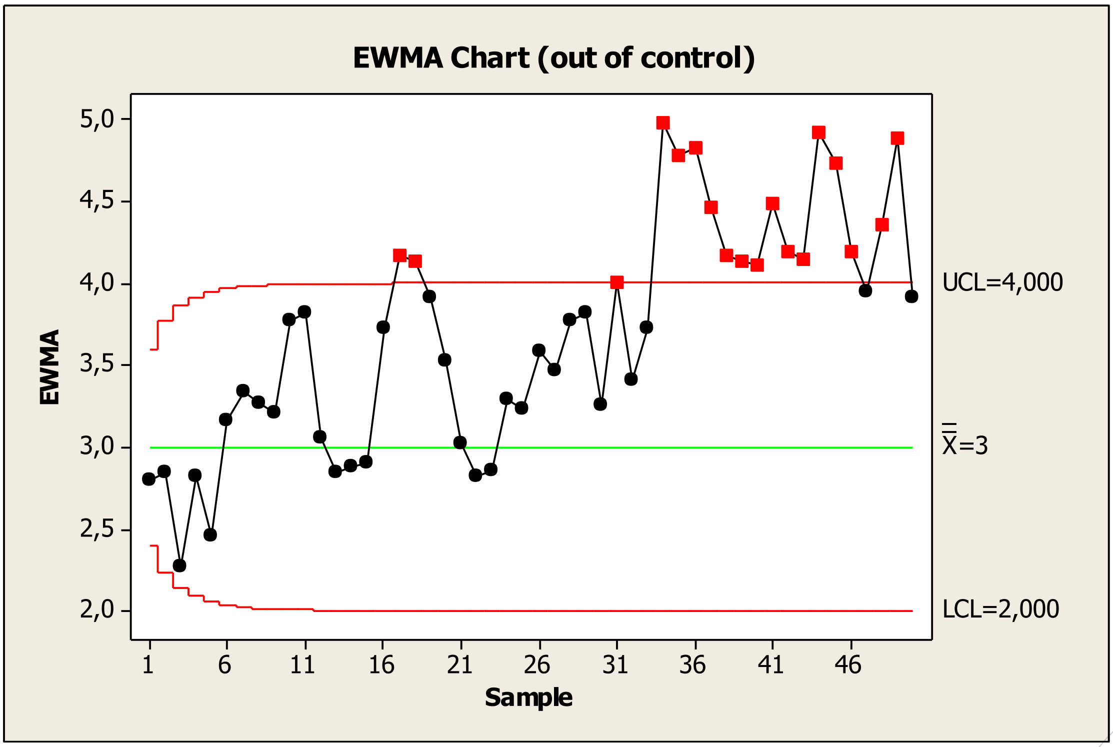

Figure 2.6. Out of control EWMA chart for simulated Poisson Data with λ=3 that changes to λ=4. The EWMA detects the change.

The EWMA Control Chart also has the ability to forecast where the process mean will for the next time period. We could say that z_i is the forecast of the value of the process mean μ at the time i+1.

2.26 The Moving Average Control Chart

The moving average control chart of span w at time i is defined as

$$M_i = \frac{x_1 + x_2 + ... + x_{i-w+1}}{w}$$

That is, at time i the oldest observation is being dropped and the newest one added to the set. The variance of the moving average M_i is

$$V(M_i)= \frac{1}{w^2} \ , \ \sum_{j=i-w+1}^{i} V(x_j) = \frac{1}{w^2}\ , \ \sum_{j=i-w+1}^{i} \sigma^2 = \frac{\sigma^2}{w}$$

The three sigma control limits are for the MA control chart are:

$$UCL=\mu_0+3\frac{\sigma}{\sqrt{w}}$$

$$CL=\mu_0$$

$$LCL=\mu_0-3\frac{\sigma}{\sqrt{w}}$$

2.27 Standardized x bar and R control charts

Standardized control charts are being recommended if the standard deviations are not the same for all parts. For the j^{th} part number we have $\overline{R_j}$ the average range and T_j the nominal value of x on this part number. The standardized R chart would be plotting the quantity: $R_i^s = \frac{R_i}{R_j}$ for all sample form this part number. The control limits would be:

$$LCL=D_3$$

$UCL = D_4$

For the standardized $\bar{x}$ chart we plot the quantity $\bar{x}_i^s = \dfrac{\overline{M}_i - T_j}{\overline{R}_j}$. The control limits would be:

$$LCL = -A2$$

$$Center\ Line = 0$$

$$UCL = A2$$

The center line is 0 because M_i is the average of the subgroups of the j th part number.

CHAPTER 3.

SHEWHART POISSON CONTROL CHARTS FOR ATTRIBUTES

3.1 Introduction

In Chapter 3 mainly Shewhart Poisson Control Charts for Attributes are dealt. More specific the C Chart for Nonconformities is being compared to the Poisson Moving Average chart and according to Khoo it seems that the Moving Average outperforms the traditional C Chart in terms of in-control and out-of-control ARLs.

In addition, the modified U Control Chart of Rudisill, Litteraland Walter is being presented. This modified chart is very effective when other sources of variability apart from the Poisson exist.

Finally, the Zero Inflated Poisson Model is introduced in order to monitor processes with excessive 0 counts. A model was developed as a zero defect process subject to random shocks. The random shock occurs with probability p and upon the occurrence of a random shock, non-conformities can be found. This number of nonconformities follows the Poisson distribution. The upper control limit of the ZIP model is also constructed based on the Jeffreys prior interval that provides good coverage probability for λ. A generalized zero inflated Poisson distribution, an extension of the ZIP model is also presented. The GZIP model is also a particular form of the Poisson distribution.

3.2 C Chart for Nonconformities Versus Poisson Moving Average

The C Chart is often used to monitor the number of nonconformities, although it can be considered slow in detecting small shifts. Khoo (2004) compares the C chart with the Poisson Moving Average Control Chart for the number of non-conformities. Comparison between the ARL of both chart types occurs.

For cases where the probability of the occurrence of a non conformity is small and samples are taken from a large inspection unit the C chart is used. A C chart is used widely in the monitoring of defect of demerit data in many industries. The C Chart is described in chapter 1.

Khoo describes the design of a Poisson Moving Average Control Chart in order to compare it to the C chart. The number of nonconformities in an inspection unit of product are c_1, c_2, ..., c_i..., and the Moving Average if width w at time i can be computed as

$$M_i = \frac{c_i + c_{i-1} + \ldots + c_{i-w+1}}{w} = \frac{\sum_{j=i-w+1}^{i} c_j}{w},\ i \geq w.$$

When $i < w$, there are not w observations to calculate a Poisson moving average of width w. For these cases the average of all observations up to period i defines the Poisson moving average at time i; i.e.,

$$M_i = \frac{\sum_{j=1}^{i} c_j}{w},\ i < w.$$

The mean of the Poisson Moving Average is:

$$E(M_i) = E(\frac{1}{w}\sum_{j-w+1}^{i} c_j) = \frac{1}{w}E(\sum_{j-w+1}^{i} c_j) = \frac{1}{w}\sum_{j-w+1}^{i} E(c_j) = \frac{1}{w}(wc) = c$$

The variance of the Poisson Moving Average is:

$$Var(M_i) = Var(\frac{1}{w}\sum_{j-w+1}^{i} c_j) = \frac{1}{w^2}Var(\sum_{j-w+1}^{i} c_j) = \frac{1}{w^2}\sum_{j-w+1}^{i} Var(c_j) = \frac{1}{w^2}(wc) = \frac{c}{w}.$$

It can be easily shown that for periods $i \geq w$ the mean and variance of the moving average are c and c/i.

Therefore the 3σ control limits for the Poisson moving average chart for periods $i \geq w$ are:

$$UCL_M = c + 3\sqrt{\frac{c}{w}}$$

$$CL_M = c$$

$$LCL_M = c - 3\sqrt{\frac{c}{w}},$$

And for i < w the limits are

$$UCL_M = c + 3\sqrt{\frac{c}{i}}$$

$$CL_M = c$$

$$LCL_M = c - 3\sqrt{\frac{c}{i}},$$

If the target value c is unavailable the limits are obtained by replacing *c* with its estimate $\bar{c}$, where

$$\bar{c} = \frac{\sum_{i=1}^{m} c_i}{m}.$$

Khoo uses two examples to compare the ARLs of the C chart and the Poisson Moving Average Control Chart. The latter outperformed the former by having lower out-of-control ARLs and higher in-control ARLs for the majority of the cases.

3.3 Modified U Charts for monitoring Poisson Attribute Processes

As mentioned in Chapter 1 the U chart is used when several inspection units are taken in the sample, increasing the area of opportunities for nonconformities. An underlying assumption of the U chart is that there are no other short term common causes of variability apart from the Poisson nature of variability exist.

When additional sources of variation exist, the proposed control limits appear too tight and there are many out-of-control signals with no detectable cause.

In order to overcome this problem the rates can be treated as continuous variable data and use an individual chart where the control limits are based on the two-point moving range. This method though ignores the Poisson effect and the dependence on the subgroup size.

In order to overcome the above mentioned problems, Rudisill, Litteral and Walter (2004) presented a method of partitioning the data into Poisson and non-Poisson sourced and using this partitioning to construct a modified U chart. They compared this to the conventional individuals chart method of dealing with the violation of the Poisson assumption. They presented a procedure that estimates the percentage of the variability due to Poisson and short-term random effects and determine control limits that reflect both sourced of common cause variability.

Many types of rate data have component of Poisson variation and an additional component of daily random non-Poisson variation. If non-Poisson variation is significant then it should be removed before the U chart can be effectively used. Usually a process engineer should identify and remove the special causes.

The method developed by Rudisill, Litteral and Walter for partitioning variability into Poisson and Non-Poisson components is presented in the following steps:

- The typical rate of occurrence should be calculated

$\bar{u} =$ Total Occurrences/Total Production Units.

- The average number if production units per subgroup (APU) should be calculated. APU=Total Production Units/Total Number of Subgroups
- The expected variance due to Poisson nature should be determined: $\sigma_u^2 = \dfrac{\bar{u}}{APU}$

- Using the actual rates, u_i (i=1,2,…,n, where n number of time periods) calculate the two point moving ranges MR_I, i=2 to n. Then calculate the average moving range, $\overline{MR}$.
- The estimate of the combined Poisson and non-Poisson variance is $\sigma_{total}^2 = \left(\frac{\overline{MR}}{1.128}\right)^2$. The degrees of fredom, df, associated with this estimate are 0.9(n-1), where n is the number of time periods (Wheeler, 1990).
- Chi-square test is performed in order to determine if the contribution due to non-Poisson sources is statistically significant. The Chi-square test statistic is $\chi^2 = df(\sigma_{total}^2)/\sigma_u^2$

a. If the calculated Chi-square value does not exceed the critical Chi-square value, conclude there is no significant non-Poisson variability.

The estimate of $\sigma_{non-Poisson}^2$ is 0.

In this case, the best chart for evaluating statistical control is the traditional U chart.

b. If the calculated Chi-square value exceeds the critical Chi-square value, conclude there is statistically significant non-Poisson variability. The estimate of the non-Poisson variance is $\sigma_{non-Poisson}^2 = \sigma_{total}^2 - \sigma_u^2$.

- Determine the relative percent contributions for the Poisson and non-Poisson sources.
- Interpret the results and focus attention on reducing the contribution for larger components.

For assessing future control, calculate control limits by:

$$UCL = \bar{u} + 3\sqrt{(\sigma_{non-Poisson}^2 + \frac{\bar{u}}{N})}$$

$$LCL = \bar{u} - 3\sqrt{(\sigma_{non-Poisson}^2 + \frac{\bar{u}}{N})}\,.$$

Following the above described steps other sources of variance can be detected and quantified and finally this variation sources can be incorporated in the construction of a modified U chart.

3.4 Attribute Control Charts for Zero-Inflated Poisson Processes

Unfortunately the traditional C chart, which is based on the Poisson distribution, has frequent false alarms when there is an excessive count of zero counts exists in a process. Due to technological advancement and automation of manufacturing processes, a well designed process could have more count of zeros than expected under its underlying Poisson distribution. This is also very common in automatic high yield manufacturing and continuous production processes. The Poisson model often underestimates the observed dispersion and therefore the control limits are improperly narrow. Another drawback of these charts is that their 3-sigma control limits evaluated based on the asymptotic normality of the attribute counts, have a systematic negative bias in their coverage probability. Hence, the basic Poisson distribution is extended so as to model larger dispersion effects. A model was developed as zero defect process subject to random shocks. The random shock occurs with probability *p* is and upon the occurrence of random shock, non-conformities can be found. This number of nonconformities follows the Poisson distribution. The distribution of the number of nonconformities is given by:

P(no nonconformities in a sample unit) = $(1-p)+pe^{-\lambda}$

and

P(k nonconformities in a sample unit) = $p\dfrac{\lambda^k e^{-\lambda}}{k!}$,

Gupta and later Bohning and Li dealt with the so called zero inflated Poisson (ZIP). It is actually a generalization of the Poisson model and as it is complicated it should only be used when the Poisson distribution is not valid. Xie, He and Goh conducted a number of comparative tests between Poisson and zero inflated Poisson distribution.

The form of the Zero Inflated Poisson distribution is the following according to Johnson and Lambert:

$$F(y;p;\mu)=\begin{cases}1-p+pe^{-\lambda}, if & y=0\\ pP_0(y,\lambda), if & y>0\end{cases}$$

With $E(Y)=p\lambda$ *and* $Var(Y)=p\lambda+p\lambda(\lambda-p\lambda)$.

The ZIP model is very easy to use and the mean and variance are of close form. The maximum likelihood estimates can also be obtained:

If $\{Y_1, Y_2,\dots,Y_n\}$ with sample size n and n_i the number of count i in the sample. The n_0 is the number of zeros in the sample.

The log-likelihood function is:

$$L(p,\lambda)=n_0 log\{1-p+pexp(-\lambda)\}+\sum_{y=1}^{\infty} n_y \log\{pP_0(y,\lambda)\}.$$

The MLE can be then obtained as:

$$p=\frac{1-n_0/n}{1-\exp(-\lambda)},\ \lambda=\overline{y}/p \text{ and } \overline{y}=\sum_{i=1}^{n} y_i/n.$$

The Hypothesis H_0: $p=1$ should be tested. If H_0 cannot be rejected then the simpler Poisson distribution should be used in it is not necessary to user the ZIP model.

Khoo presents a number of tests used to test the Poisson model against the ZIP model: The score test of Vandebroek, the likelihood ration test, the Chi-Square Test, the Cochran test, the Rao-Chajravarti test and a test based on a confidence interval of p.

When the count data can be modeled by a ZIP model, statistical process control procedures can be modified. The lower control limit for the ZIP model will not exist which is common for attribute chart. The upper control limit UCL-n_u for the control chart based on the number of nonconformities can be obtained as the smallest integer solution of

$P(n_u$ or more nonconformities in a sample$) \leq \alpha_L$,

where α_L is the predetermined false alarm probability for the UCL n_u.

In conclusion a ZIP model can be used as an alternative to the traditional Poisson model, especially when the data shows over-dispersion with the traditional Poisson. In that way the drawback of having many false alarms and frequent stopping of the procedure will be overcome, since more appropriate upper control limit can be derived.

Sim and Lim suggested fitting zero inflated models to the zero inflated attribute count and then replace the 3-sigma control limits with limits constructed based on a confidence interval that provides good coverage probability for the parameter under study.

More specific, they constructed a novel one-sided c_j-chart that can be used to monitor the zero-inflated Poisson count. This chart is based on the estimated value of Poisson parameter λ in the ZIP Model and an upper one sided Jeffreys prior interval for λ.

The proposed chart is also constructed using a simple two-of-two control rule to enhance its performance in detecting upward shifts of the process parameter under study. As it is impossible to tell which of the observed zeros are due to chance variation allowed for under the presumed distribution and which are the excess zeros, thus only the one-sided attribute charts are constructed to detect upward shifts of the process parameter using the positive attribute count.

Sim and Lim constructed the C_j chart as the classical one-sided c-chart for the Zero Inflated Poisson distribution does lead to a smaller false alarm rate but on the other hand it does not yield the false alarm rate and the desired ARL until a false alarm. The reason for this is the poor coverage probability of the upper 3 –sigma control limit used in the C chart.

Xie constructed a one sided C chart based on the ZIP model to detect upward shifts so as to overcome the above mentioned problem. This chart was constructed with its UCL based on the assumption that increases in the values of the parameters p and λ would lead to large nonconformities count. However, the disadvantage of this chart is that increases in p would lead to increasing number of none-zero Poisson counts, but not necessarily to larger nonconformities counts.

Sim and Lim suggested the alternative C_j chart which is based on the estimated value of Poisson parameter λ in the ZIP Model and an upper one sided Jeffreys prior interval for λ. If X the number of nonconformities that follow a Poisson (λ) distribution. If $X=k$ then the (1-a)100% Jeffreys prior interval is defined as:

$$CI_J^{\lambda}(k) = [G(\alpha); k+0.5,1), \infty)$$

with $CI_J^{\lambda}(0) = [0, \infty)$, where $G(\alpha;a,b)$ denotes the 100$^{\text{th}}$ percentile of a Gamma distribution with shape parameter a and scale parameter b. Note that the value $G(\alpha;x+0.5,1)$ increases when the value of x increases, hence the length of the Jeffreys interval $CI_J^{\lambda}(k)$ decreases as x increases. Cai (2005) pointed out that the Jeffrys prior interval yields coverage probability that is closer to the desired probability than the confidence interval obtained under the normal approximation.

The C_j chart is displayed by plotting the inspection unit number against the observed nonconformities Poisson count k in each of the successive inspection units. The upper control limit of the proposed C_j -chart is defined as the largest Poisson count k such that the estimated value, λ_0, of λ estimated from falls inside the (1-a) 100% interval $CI_J^{\lambda}(k)$ ΄that is:

$$UCL_{c_j}(\lambda_0) = \max[\, x \mid \lambda_0 > G(\alpha; x+0.5,1)]$$

A current or future Poisson count that plots above the $UCL_{c_j}(\lambda_0)$ is then interpreted as evidence that the process is out-of-control with an upward shift of λ from λ_0 to a larger value.

Chen, Zhou, Chang and Huang proposed another very flexible distribution the generalized ZIP (GZIP) for monitoring attribute data. It is actually an extension of the ZIP model. It is also a particular form of compound Poisson distributions. The GZIP distribution can be used to monitor various attribute data.

GZIP distribution is an extension of the above discussed ZIP distribution. In real-world applications the processes experience different kind of shocks, which have different influences on the process and therefore result in different numbers of nonconformities. If the number of nonconformities caused by each different shock follows a specific Poisson distribution with parameter λ_i and the occurring probability if each shock is μ_i then the distribution of the number of nonconformities in the process with n kinds of shocks can be expressed as

$$P(X=0)=1-\sum_{i=1}^{n}\mu_i+\sum_{i=1}^{n}\mu_i e^{-\lambda_i}$$

$$P(X=k)=1-\sum_{i=1}^{n}\mu_i\frac{\lambda_i^x e^{-\lambda_i}}{k!}$$, $(x>0)$ or

$$P(X=k)=(1-\sum_{i=1}^{n}\mu_i)\cdot I(X=0)+\sum_{i=1}^{n}\mu_i\frac{\lambda_i^x e^{-\lambda_i}}{k!},$$

Where *I(.)* is the indicator function, which equals 1 when the condition is true and equals 0 otherwise. When $n=1$, GZIP degenerates into the original ZIP model. Obviously 2n parameters should be determined when the number of shock types is n. Therefore a key problem for using the GZIP Poisson distribution would be to determine all parameters that represent the in-control data. Mathematically, it can be stated as: given a series of independent random samples $\{X j \mid j = 1,2, \ldots, N\}$, estimate the potential number of shock types n, and corresponding parameter λ_i μ_i ,which can best approximate the true distribution of the data.

Chen, Zhou, Chang and Huang used the Expectation-Maximization (EM) algorithm to estimate the parameters of the model.

GZIP distribution can be used to construct several control charts to-monitor various types of counting data with great flexibility: Similar to other types of Shewhart charts on attribute data, the statistic that GZIP chart monitors is the present sample of counting data. Given the specified x-error, and the estimated parameter $\boldsymbol{\theta}$ from historical in-control data, we can calculate the upper control limit accordingly (the lower control limit can be set to zero) as $P(x>\text{UCL}|\boldsymbol{\theta}) \leq \alpha$. Therefore, when using this chart for monitoring, we will generate alarms whenever the sample value exceeds the upper control limit. The theoretical average run length (ARL) for in-control data is $1/\alpha$.

CUSUM and Ranked probability control charts can also be designed for the Poisson distribution.

The GZIP model has still some open issues: The noises, that real data always contain, have very large impacts on the estimated model if the model is sensitive to it. The sample size determination is another issue. It is interesting to determine how many samples are enough to estimate the real distribution.

CHAPTER 4.

POISSON EWMA AND CUSUM CONTROL CHARTS

4.1 Introduction

In this section several EWMA as well as CUSUM control charts are being examined. Apart from the Poisson EWMA control chart the Double EWMA chart is being presented, which requires a second Exponentially Weighted Moving Average, hence the D which stands for double. Moreover, the Generally Weighted Moving Average Control Chart developed by Sheu and Lin (2008) is presented. This chart had been developed to monitor Poisson observations and detect persistent, small process shifts. The PGWMA is actually a generalized charting model for which the PEWMA chart and c-chart are special cases. In addition other PGWMA control charts are being presented.

As far as the Poisson CUSUM charts are concerned the conditional and marginal performance of a Poisson CUSUM chart when the parameter is unknown is examined. Finally the CUSUM control chart based on the Poisson distribution compounded by a Geometric distribution by Chen Randolph and Liou is presented.

4.2 Poisson EWMA Control Chart

Borror, Champ and Rigdon (1998) studied the Poisson EWMA control chart mentioned in the previous chapter: The exponentially weighted moving average is: $z_i = \lambda x_i + (1-\lambda) z_{i-1}$ with $z_0 = \mu_0$, and the control chart parameters are as follows:

$$\text{UCL} = \mu_0 + A_u \sqrt{\frac{\lambda \mu_0}{2-\lambda}[1-(1-\lambda)^{2i}]}$$

Center line=μ_0

$$\mathrm{LCL} = \mu_0 + A_l \sqrt{\frac{\lambda\mu_0}{2-\lambda}[1-(1-\lambda)^{2i}]}$$

A_u and A_l are the upper and lower control limit factors. We use often $A_l = A_u = A$.

The above limits are based on the exact variance: $Var(z_i) = \frac{\lambda\mu_0}{2-\lambda}[1-(1-\lambda)^{2i}]$, but they can also be based in the asymptotic variance $Var(z_t) \approx Var(z_\infty) = \frac{\lambda\mu_0}{2-\lambda}$.

Statistical performance of Control Charts is usually evaluated using average run lengths and standard deviations of run lengths. Borror, Champ and Rigdon used the Markov chain approach to obtain the ARL_1 for the Poisson EWMA Control Chart (using the asymptotic variance) and the results showed that the ARL_1 is usually smaller than that of the Shewhart c chart, which can also be used for Poisson data or the three EWMA charts proposed by Gan (1990) (REWMA, CEWMA, EWMA).

Moreover, the PEWMA Control Chart has usually a positive control limit, thus downward changes in the mean can also be detected.

4.3 Poisson DEWMA Control Chart

Zhang, Govindaraju, Lai and Bebbington (2003) introduced the Double Exponentially Weighted Moving Average Control Chart for monitoring Poisson Data. This requires a second Exponentially Weighted Moving Average, hence the D which stands for double.

If x_1, x_2, x_3, x_4,…,x_i…i.i.d. Poisson random variables with mean μ. If $\mu=\mu_0$ where μ_0 is known, then the process is in control, otherwise the process is out of control. To monitor the changes in the process mean the DEWMA control statistic y_i is defined via the system of equations: $z_i = \lambda x_i + (1-\lambda) z_{i-1}$, $i \geq 1$,

$z_0 = \mu_0$,

$y_i = \lambda z_i + (1-\lambda) y_{i-1}$, i≥1,

$y_0 = \mu_0$,

where 0<λ<1 is a smoothing constant. It is obvious that z_i is the usual EWMA control statistic and y_i, the DEWMA control statistic, is an EWMA of z_i.

The variance of y_i is:

$$Var(y_i) = \mu_0 \frac{\lambda^4 (1 + (1-\lambda)^2 - (i+1)^2 (1-\lambda)^{2i} + (2i^2 + 2i - 1)(1-\lambda)^{2i+2} - i^2 (1-\lambda)^{2i+4}}{[1-(1-\lambda)^2]^3}.$$

The PDEWMA chart is constructed by plotting z_i against i for i=1, 2,..., with the control limits set as

$$UCL = \mu_0 + K\sqrt{Var(y_i)} \qquad \text{Center line} = \mu_0$$

$$LCL = \mu_0 - K\sqrt{Var(y_i)},$$

Where K>0 a control limit constant which, together with the smoothing constant λ determines the performance of the PDEWMA control chart. Time-varying Variance (and not asymptotic) control limits are being used.

In order to evaluate the PDEWMA control chart Zhang, Govindaraju, Lai and Bebbington compare its ARLs and SDRLs with those of the PEWMA control chart.

They consider the smoothing constants λ=0.05, 0.1, 0.2, 0.25, 0.3, 0.4 and 0.5 for the PDEWMA chart and the PEWMA and consider distributions with μ_0=4, 8, 12 or 20. Then they list the pairs (λ, K) (λ, A) that give an in control ARL of about 200 for the PDEWMA and the EWMA respectively. The values for the PEWMA are found using a Markov Chain Approach, whereas the values for the PDEWMA are obtained through simulation.

The authors of the article followed the steps below for the simulation:

1. Random Poisson numbers were generated.
2. The control statistics z_i and y_i were calculated.
3. The control statistic is compared with an experimental UCL and LCL and the corresponding run lengths are recorded.
4. After 10,000 simulation runs and 10,000 recorded run lengths the sample mean and square root of sample variance give the ARL and SDRL respectively.
5. The search stops for A and L when the difference of the derived and target ARL is greater than 0 but less than 0.5.

Then, for various different sample means Poisson random data were simulated and ARLs and SDRLs for PEWMA and PDEWMA were determined. The simulation to derive these is from step 1 to step 4 (since A and L had already been determined and so the control limits are preset).

From the tables obtained from this simulation the conclusion is that the PDEWMA control chart is improves upon the PEWMA control chart, since

1. All out of control ARLs of the PDEWMA are smaller than those of the PEWMA control chart with a more significant difference when the process mean change is small.
2. The out-of-control SDRLs of the PDEWMA are smaller than those of the PEWMA chart.
3. The in-control SDRLs of the PDEWMA are approximately equal to those of the PEWMA chart, except from the cases were λ=0.05 and 0.1.

It should be mentioned, that when the lower control limit of the PEWMA is positive (which is usually the case) the lower limit of the PDEWMA is also positive.

Finally, the authors recommend the use of λ in the range [0.1, 0.5] since for very small λ (e.g. 0.05) the in-control SDRLs of the PDEWMA are larger to those of the PEWMA chart.

4.4 Poisson GWMA Control Chart

The Generally Weighted Moving Average Control Chart developed by Sheu and Lin (2008) had been developed to monitor Poisson observations and detect persistent, small process shifts. The PGWMA is actually a generalized charting model for which the PEWMA chart and c-chart are special cases. As the PEWMA and the PDEWMA chart, its lower control limit is usually positive so that the downward shift of a process mean can also be detected.

Sheu and Chiu (2007) show based on simulation results that the statistical performance of the best performing PGWMA chart is superior to the PEWMA chart and the c chart.

The GWMA Control Chart is a generalized weighted moving average of sequential historical data, in which each datum is assigned a different weight that decreases from the present period to the remotely past period such that the GWMA can reflect the important information on recent process.

The GWMA statistic at the jth time period can be written as

$$Y_i = \sum_{k=1}^{i} \Pr(M = k) x_{i-k+1} + \Pr(M > i) Y_0 = \sum_{k=1}^{i} (\overline{P}_{k-1} - \overline{P}_k) x_{i-k+1} + \overline{P}_i Y_0$$

Where $Y_0 = \mu_0$ the mean of the process,

the observation x_1, x_2, .. are i.i.d.,

and M the number of samples until the first occurrence of the event for which the control statistic goes beyond control limits.

It also holds that $\Pr(M = k) = (\overline{P}_{k-1} - \overline{P}_k)$.

The weighting factor $\overline{P}_i = \Pr(M > i) = q^{i^a}$, where 0 ≤ q ≤ 1, α>0, i=1, 2, (Note that q is a constant and α is determined by the practitioner).

The control statistic Y_i is then determined as:

$Y_i = \sum_{k=1}^{i}(q^{(k-1)^a} - q^{k^a})x_{i-k+1} + q^{i^q}\mu_0$, where 0 ≤ q ≤ 1, α>0, i=1, 2, …

The mean and variance of the control statistic Y_i are μ_0 and $Q_i\sigma^2$ respectively

Where $Q_i = \sum_{k=1}^{i}(q^{(k-1)^a} - q^{k^a})^2$, for every i=1,2,...

The GWMA control chart can be constructed as

$$UCL = \mu_0 + L\sqrt{Q_i}\sigma$$

Center line=μ_0

$$LCL = \mu_0 - L\sqrt{Q_i}\sigma,$$

Where L the constant to match the desired ARL for a specific GWMA control chart.

It should be noted that for a=1 and λ=1-q the GWMA control chart transforms to the simple EWMA control chart. For the special case of q=0 we obtain the Shewhart control chart.

Sheu and Chiu determine appropriate parameters with simulation and develop the PGWMA control chart to measure performance with ARL.

For the Poisson distribution the control limits can be also written as:

$$UCL = \mu_0 + L\sqrt{Q_i\mu_0}$$

Center line=μ_0

$$LCL = \max\{0, \mu_0 - L\sqrt{Q_i\mu_0}\},$$

since Poisson numbers are non negative and the variance equals the mean of the distribution.

In order to design the PGWMA control chart following steps are followed:

1. Poisson random numbers are generated and $Y_{i,}$ the PGWMA control statistic, is calculated.
2. The control limits are calculated given a appropriate value for L.
3. After 40,000 iterations ARL and SDRL are calculated
4. Using the idea of ‘inverse regression’ and repetition of the steps 1-3 ARL_0 is obtained for the in-control process
5. Under the (q, a, L) combination for the desired ARL_0 the ARLs and SDRLs for given shifts can be calculated by performing steps 1-3.

The conclusion of Sheu and Chiu is that by adding an adjustment parameter to the PEWMA model three advantages are obtained:

1. Especially for very small process mean shifts it rapidly detects the out-of-control signal
2. The Shewhart c-chart and the PEWMA are special cases of the PGWMA control chart
3. As for the PEWMA and PDEWMA control chart its lower control limit is usually positive so that downward shifts can also be detected.

In 2008 Chiu and Sheu introduce different PGWMA control charts: ACLPGW, a PGWMA chart with asymptotic control limits,

VCLPGW, a PGWMA chart with time-varying control limits,

FVCLPGW, a FIR (Fast Initial Response) PGWMA chart with time-varying control limits

FADJPGW, a FIR (Fast Initial Response) PGWMA chart with time-varying control limits

and

PDGWMA a Poisson Double GWMA control chart

All these different variations of the GWMA control chart are presented in the following table:

		Control Limits	
		Time Varying	**Asymptotic**
Fast initial Response Device	**No**	VCLPGW	ACLPGW
		PDGWMA	
	Yes	FVCLPGW	Not Studied
		FADJPGW	

The VCLPGW Control Chart has already been described (Chiu and Sheu 2007).

The FVCLPGW Control Chart has a two-sided 50% head – start as follows:

$$Y_0^U = \mu_0 + 0.5L\sqrt{Q_1\mu_0}$$

$$Y_0^U = \mu_0 + 0.5L\sqrt{Q_1\mu_0} \ ,$$

Where $Q_1 = (1-q)^2$ and for i=1,2, it is formulated as follows:

$$Y_i^U = \sum_{k=1}^{i}(q^{(k-1)^a} - q^{k^a})x_{i-k+1} + q^{i^q}Y_0^U$$

$$Y_i^L = \sum_{k=1}^{i}(q^{(k-1)^a} - q^{k^a})x_{i-k+1} + q^{i^q}Y_0^L \ ,$$ for every i=1, 2,...

The FADJPGW uses an exponentially decreasing adjustment method to further narrow the limits. Its control limits are as those of the VCLPGW Control Chart but adjusted as follows:

$$UCL_i^{adj} = \mu_0 + F_{adj}\sqrt{Q_i\mu_0}$$

Center line=μ_0

$$LCL_i^{adj} = \max\{0, \mu_0 - F_{adj}\sqrt{Q_i\mu_0} \ ,$$

Where F_{adj} denotes the FIR adjustment factor:

$$F_{adj} = 1-(1-f)^{1+a(i-1)},\ 0<f\leq 1,\ a>0,\ i=1,2,\ldots$$

It should be noted that a should be set so that the FIR adjustment has a very little effect after observation 20.

The novel control chart introduced in Chiu and Sheu's paper is the PDGWMA Control Chart.

This Poisson Double Generally Weighted Moving Average Control Chart doubly smooths the Poisson observation. The PDGWMA Control Chart is defined by:

$$Y_i = \sum_{k=1}^{i} p_k X_{t-k+1} + q^{i^a} Y_0$$

$$Z_i = \sum_{k=1}^{i} p_k Y_{t-k+1} + q^{i^a} Z_0 \text{ , i=1,2,...}$$

Where Y_0 and Z_0 the target mean μ_0 and p_k the weighting factor

$$p_k = q^{(k-1)^a} - q^{k^q} .$$

Z_i can also be written as:

$$Z_i = \sum_{k=1}^{i} w_k X_{i-k+1} + (1-\sum_{k=1}^{i} w_k)\mu_0 \text{ , i=1,2,...}$$

Where $w_k = \sum_{j=1}^{k} p_j p_{k-j+1}$.

The mean and variance of the control statistic are $E(Z_t)=\mu_0$ and $Var(Z_t)=(\sum_{j=1}^{i} w_j^2)\sigma^2$ for i=1,2,...

The PDGWMA control chart can be designed as:

$$UCL_i^D = \mu_0 + L\sqrt{Var(Z_i)}$$

Center line=μ_0

$$LCL_i^D = \max\{0, \mu_0 - L\sqrt{Var(Z_i)}\},$$

Using simulation parameter combinations (q,a,L) for all different charts with common ARL_0 are obtained and for various shifts the ARLs and SDRLs are obtained for comparison.

Chiu and Sheu conclude the following after the comparison of these control charts.

1. PWMA Control Charts with FIR feature are useful when detecting great shifts in the initial stage.
2. ACLPGW, for which no FIR feature is provided, is the worst in detecting start-up quality problems.
3. PDGWMA is the first choice, especially for downward shifts
4. FADJPGW is the best in detecting great process shifts $\geq 1.5\sigma$ because its control limits are further narrowed, in order to improve the FIR feature.
5. FVCLPGW chart is better than FADJPGW in detecting small shifts <1s.
6. VCLPGW is very good in detecting small upward shifts.

4.5 Poisson CUSUM Control Chart-Conditional and Marginal Performance

The CUSUM control chart is also appropriate to monitor Poisson data. Assume x_1, x_2,…are i.i.d. Poisson random variables with mean μ. The process is in control when $\mu = \mu_0$, where μ_0 the value of the in control process mean. The process is out of control when $\mu \neq \mu_0$.

If the mean μ_0 is unknown, a Phase I study should be conducted to obtain from an in-control estimated sample mean $\hat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n} x_i$ to use instead of μ_0.

In his paper Testik (2007) examines the conditional and marginal performance of a Poisson CUSUM chart when the parameters (in case of a Poisson distribution the parameter) are unknown. For the Poisson distribution K is determined to distinguish between an acceptable mean value μ_a and a detectable mean value μ_d that the CUSUM is to detect quickly. The reference value for the Poisson CUSUM is given by:

$$K = \frac{\mu_d - \mu_a}{\ln(\mu_d) - \ln(\mu_a)}.$$

Generally μ_a is chosen to be the current system performance (μ_0 or estimated $\hat{\mu}_0$). The μ_d represents the out-of control mean for which the CUSUM is being set to determine quickly. It should be noted at this point, that unlike the normal distribution where K lies between in-control and out-of control mean values for the Poisson case K is lying closer to μ_a.

The Poisson CUSUM is determined by

$$C_i^+ = \max[0, x_i - K + C_{i-1}^+]$$
$$C_i^- = \max[0, K - x_i + C_{i-1}^-]$$

Where the starting values are $C_0^+ = C_0^- = 0$

The decision interval H for the CUSUM is determined in combination with K on the basis of ARLs. As for all charts H is selected to give large ARL_0 and a small ARL_1.

Testik mentions, that when the mean of the process is unknown μ_a is set to be equal to the estimate $\hat{\mu}_0$ for determining the K and finally select H in combination with K (usually through a computer program or a table look-up procedure). For that reason for different ARL_0 and different estimated means $\hat{\mu}_0$, there are different pairs of K and H. Thus, K and H can be viewed as random variables depending on the estimated mean of the process $\hat{\mu}_0$ and can be noted as $\hat{K}$ and $\hat{H}$ respectively. For this reason the performance of a Poisson CUSUM actually depends on the accuracy of estimating μ_0.

For the purpose of his paper Testik investigates only the CUSUM for increases in the mean number of counts. $\hat{K}$ is rounded to integer values and $\hat{H}$ is selected to be a positive integer.

The Markov chain approach is being extended for computing the conditional and marginal performance of Poisson CUSUM when the estimated in-control mean $\hat{\mu}_0$ differs from the actual mean of the process μ_0. The conditional performance measures ARL and SDRL given $\hat{\mu}_0$, $\hat{K}$, $\hat{H}$ and C_0^+. The actual parameters are assumed in that case different and are hypothetically known for evaluation. The marginal performance can be calculated by integrating the conditional performance measures over the distribution of the estimated in-control mean. As for the marginal performance, evaluation does not require knowledge of the specific parameters and for that reason it is useful for drawing conclusions regarding the sample size requirements.

Conclusively, Testik proves with several examples that a Poisson CUSUM Control Chart constructed using parameter estimates may perform poorer than the chart constructed with known parameters. Especially the effect on the in-control Run Length may be significant, if the mean estimate is not good enough. If the sample size is greater than or equal to 200 this would result in a Poisson CUSUM Control Chart with and in an out of control performance close to that of a Chart with a known process mean.

4.6 CUSUM Control Chart for the Compound Poisson Distribution

Chen, Randolph and Liou presented a CUSUM control chart based on the Poisson distribution compounded by a Geometric distribution. The assumption made that the defects of a process follow a Poisson distribution is not always valid, since the process is more complex in practice and the distributions of defects are more appropriately modeled by the compound Poisson distribution. An effective CUSUM control scheme can be obtained from the probability transition matrix for the Markov chain proposed by Brook and Evans (1972).

Chen, Randolph and Liou mention the example of automobile accidents on a given street. The number of accidents follows a Poisson distribution, however the number of injuries that can occur for each accident cannot be described by a Poisson. More precisely, the number of injuries in a given period is said to have a compound Poisson distribution. Similarly, the common assumption for the number of defects per unit follows the Poisson distribution. In some situations, however, the defects generating process is more complex and the distributions of defects are more appropriately modeled by the compound Poisson distribution.

Even a small shift can be costly if it is not detected quickly. Since small shifts are not easily detected in small periods of time any improvement as far as the fast detection is concerned would be very beneficial. CUSUM control charts are considered to be very sensitive to small sustained shifts in a process. With current and previous process information, the CUSUM control scheme is capable of detecting out-of-control situations caused by small sustained changes or shifts in the process mean quicker. Also, it may locate the time of change effectively. CUSUM techniques are widely used for quality control and process monitoring in companies, such as Du Pont and Motorola's Semiconductor Process Sector (Lucas, 1985; White et al., 1997).

For the CUSUM Control Chart the distribution is usually assumed to be Poisson. A defective item can have one or multiple defects. The number of defects per item does not follow the Poisson but another distribution, the compound Poisson distribution. The compound Poisson distribution of defects is reported in IC fabrication
by Stapper (1985), Gardiner (1989), and Albin and Freidman (1989). Meanwhile, Randolph and Sahinoglu (1995) reported the application of the geometric Poisson distribution, a member of the compound Poisson family, for software quality control. In this paper, we take one step further to develop the CUSUM control scheme for the geometric Poisson production process to detect a small sustained shift in the mean.

Chen, Randolph and Liou obtain the ARLs from the probability transition Matrix for the Markov chain proposed by Brook and Evans (1992). Then they compare the ARLs between the geometric Poisson and the Poisson CUSUM control schemes.

A compound distribution is formed by mixtures of discrete distribution. The result of this 'mixing' of distributions is called compound distribution. The symbol for these distributions is:

$$F_1 \underset{\theta}{\wedge} F_2,$$

Where F_1 represents the original distribution, F_2 the compounding and θ the varying parameter. In our case F_1 is the Poisson distribution and the compound distribution is called the Poisson distribution compounded by F_2 and denoted:

$$Poisson \underset{\theta}{\wedge} F_2$$

In many examples in which events occur according to the Poisson distribution, and, furthermore, for each of these Poisson events one or more other events, say E_s, can occur. In this situation, the process of generating events E_s is said to follow the compound Poisson process. An example given by Sherbrooke (1966), describing the items demanded process in the inventory system of a store. The focus was on the items demanded by customers. Customer arrivals were assumed, in general, to follow the Poisson distribution, and each customer could demand a positive discrete amount. Thus, the number of demands in a time interval is said to have a compound Poisson distribution. Also, Sherbrooke (1966) stated that the compounding distribution of the number of demands by one customer is geometric; and thus, the number of demands in a time interval is the geometric Poisson distribution.

The authors use the term 'defective' item to represent the Poisson event and 'defect' to represent the compounding event. In a unit there are λ defective items. Each item contains at least one defect and another defective item would be occurring with probability ρ. For the Poisson Distribution the time between defective items is independent of the time at which earlier defective items where produced.

If $Y_{(t)}$ the random variable of the number of defective items and $X_{(t)}$ the random variable of the number of defects that occur up to t, where $t > 0$. The density function of the geometric Poisson with parameter λ and ρ can be written as:

$$P[(X(t))=0]=e^{-\lambda t}$$

$$P[(X(t))=x]=\sum_{y=1}^{x}\frac{(\lambda t)^{y}e^{-\lambda t}}{y!}\times\binom{x-1}{y-1}\rho^{x-y}(1-\rho)^{y}\text{ , x=1,2,...}$$

Where $\lambda>0, 0<\rho<1$.

Based on the above density function described above the expected number of defects for each fixed unit t can be calculated. For *t=1* the expected value of *X* is:

$$E(X)=\sum_{x=1}^{\infty}x\sum_{y=1}^{x}\frac{\lambda^{y}e^{-\lambda}}{y!}\binom{x-1}{y-1}\rho^{x-y}(1-\rho)^{y}$$

$$=\sum_{y=1}^{\infty}\sum_{x=y}^{\infty}x\frac{\lambda^{y}e^{-\lambda}}{y!}\binom{x-1}{y-1}\rho^{x-y}(1-\rho)^{y}$$

$$=\sum_{y=1}^{\infty}\frac{\lambda^{y}e^{-\lambda}}{y!}\sum_{x=y}^{\infty}x\binom{x-1}{y-1}\rho^{x-y}(1-\rho)^{y}$$

$$=\sum_{y=1}^{\infty}\frac{\lambda^{y}e^{-\lambda}}{y!}y\frac{1}{1-\rho}\sum_{x=y}^{\infty}\binom{x}{y}\rho^{x-y}(1-\rho)^{y+1}$$

The second summation is the probability mass function of a negative binomial distribution with parameters *x+1* and *y+1* and therefore:

$$E(X)==\sum_{y=1}^{\infty}\frac{\lambda^{y}e^{-\lambda}}{y!}y\frac{1}{1-\rho}=\frac{1}{1-\rho}\sum_{y=1}^{\infty}\frac{\lambda^{y}e^{-\lambda}}{y!}y=\frac{\lambda}{1-\rho}.$$

The variance of X can be derived as described below:

$$E[(X+1)X]=\sum_{x=1}^{\infty}(x+1)x\left[\sum_{y=1}^{x}\frac{\lambda^{y}e^{-\lambda}}{y!}\binom{x-1}{y-1}\rho^{x-y}(1-\rho)^{y}\right]$$

$$=\sum_{y=1}^{\infty}\sum_{x=y}^{\infty}(x+1)x\cdot\frac{\lambda^{y}e^{-\lambda}}{y!}\binom{x-1}{y-1}\rho^{x-y}(1-\rho)^{y}$$

$$=\sum_{y=1}^{\infty}\frac{\lambda^{y}e^{-\lambda}}{y!}\sum_{x=y}^{\infty}(x+1)x\cdot\binom{x-1}{y-1}\rho^{x-y}(1-\rho)^{y}$$

$$=\sum_{y=1}^{\infty}\frac{\lambda^{y}e^{-\lambda}}{y!}\sum_{x=y}^{\infty}(x+1)x\cdot\frac{(x-1)!}{(x-y)!(y-1)!}\rho^{x-y}(1-\rho)^{y}$$

$$=\sum_{y=1}^{\infty}\frac{\lambda^{y}e^{-\lambda}y(y+1)}{y!(1-\rho)^{2}}\sum_{x=y}^{\infty}\cdot\frac{(x+1)!}{(x-y)!(y+1)!}\rho^{x-y}(1-\rho)^{y+2}$$

$$= \frac{1}{(1-\rho)^2}\left(\sum_{y=1}^{\infty} y^2 \cdot \frac{\lambda^y e^{-\lambda}}{y!} + \sum_{y=1}^{\infty} y \cdot \frac{\lambda^y e^{-\lambda}}{y!}\right)$$

$$= \frac{1}{(1-\rho)^2}[E(Y^2)+E(Y)]$$

If Y is distributed according to the Poisson distribution with parameter λ, then

$$E(Y^2)=Var(Y)+[E(Y)^2]=\lambda+\lambda^2$$

So,

$$E[(X+1)X] = \frac{1}{(1-\rho)^2}[2\lambda+\lambda^2]$$

Since $E(X^2)=E[X(X+1)-E(X)]$, *then*

$$Var(X)=E(X^2)-[(E(X)]^2=E[(X+1)X]-E(X)-[E(X)]^2$$

$$\frac{2\lambda+\lambda^2}{(1-\rho)^2} - \frac{\lambda}{1-\rho} - \left(\frac{\lambda}{1-\rho}\right)^2 = \frac{\lambda(1+\rho)}{(1-\rho)^2}.$$

It is obvious that the variance of the geometric Poisson distribution is greater or equal to the mean. If the variance equals to the mean the geometric Poisson equals to the Poisson. Sherbrooke (1966) also shows that the compound Poisson distribution is ‘memoryless’; that is, the number of failures occurring in a fixed unit t does not influence the probabilities of failures in any other unit. Feller (1950) proves that the compound Poisson is the most general class of discrete distributions that are memoryless. The geometric Poisson distribution has been studied and applied by Randolph and Sahinoglu (1995) to problems of control of defects in software. Other applications of the compound Poisson can be found in Barbour, Chryssaphinou, and Vaggelatou (2001).

Chen, Randolph and Liou describe then the necessary steps to implement a geometric Poisson CUSUM. The steps are almost the same as for a normal CUSUM:

1. Firstly, the acceptable and the non acceptable process mean must be determined.
2. The smallest acceptable in-control ARL (ARL_0) are then selected.
3. Select possible *Ks* between the acceptable and the non acceptable mean.
4. Based on *K* and *H* ARL_0 should be calculated for pilot study. Several combinations for ARLs which are greater or equal to the smallest ARL_0 should be selected.
5. Compare the out-of-control ARL (denote ARL1) forthe (H,K) combination with other choices of (H,K) producing the same in-control ARL. The CUSUM scheme which gives the most desired performance interms of out-of-control ARL is selected.

The authors also propose a procedure that can help the practitioner determine if the underlying distribution is the geometric Poisson. More specific, the steps are the following:

1. A number of *n* sample should be selected and the number of defects in each sample must be denoted $X_1, X_2, \dots, X_n$.
2. The estimates of *E(X)* and *Var(X)* for geometric Poisson random variables must be calculated. Given those unbiased estimators of the mean and Variance the equations:

$\overline{X} = \frac{\hat{\lambda}}{1-\hat{\rho}}$ and $S^2 = \frac{\hat{\lambda}(1+\hat{\rho})}{(1-\hat{\rho})^2}$ should be solved to obtain:

$\hat{\lambda} = \frac{2\overline{X}^2}{S^2 + \overline{X}}$ and $\hat{\rho} = \frac{S^2 - \overline{X}}{S^2 + \overline{X}}$, which are estimates of the parameters λ and ρ.

3. Observed and expected frequencies must be summarized in each of *n* sample and *c* classes.
4. The following Hypotheses:

H_0: The form of the distribution of defects is geometric Poisson.

H_1: The form of the distribution of defects is not geometric Poisson.

must be tested. The Hypotheses can be tested using the Chi-Squared Test.

CHAPTER 5.

CONTROL CHARTS FOR MULTIVARIATE POISSON DISTRIBUTION

5.1 Introduction

Current literature suggests many methods for monitoring multivariate means whose populations are assumed to be normally distributed. A model-based control chart for a multivariate Poisson with measurable inputs had been proposed by Skinner et al. (2003). By assuming the independence between output variables, the deviance residuals resulting in generalized linear model were used in the monitoring scheme for each variable. Chiu and Kuo introduced the multivariate Poisson chart in order to monitor the correlated multivariate Poisson count data.

5.2 The multivariate Poisson Chart

The charts that had been examined by Lowery and Montgomery in order to monitor data that follow the Multivariate Poisson distribution are the Hotelling multivariate control chart, multivariate cumulative sum (MCUSUM) and the multivariate exponentially weighted moving average control chart (MEWMA). Bersimis et al. discussed the application of principal components and partial least squares in the construction of the multivariate control chart. However, in praxis, where multiple correlated counts data occur other type of charts are needed. The quality characteristics of multiple attributes process would follow either a multiple Binomial distribution or multivariate Poisson distribution. When the focus is on the number of defects on each unit and the number of defects is classified into more than two categories, the data could be modeled using the multivariate Poisson distribution.

Patel developed a quality control method similar to the Hotelling type, however this was a very complex and therefore not widely used method. Another approach is the np control chart established by Lu et al in order to deal with the multivariate attribute process. They also proposed an approach to identify which quality characteristic was the major contributor to an out-of control signal. However the normality of the process is an assumption that is not always valid. Jiang et al. (2002) proposed a symmetric c-chart and obtained the control limits to minimize the absolute deviation of $ARL_{0.}$ It solved the problem that the Shewhart-type 3-sigma control limits with normal approximation may result in a large deviation of ARL_0 λ the Poisson parameter is small. The methods for finding optimal control limits can also be applied to monitor multivariate Poisson counts with positive correlation. However, the discussion on out-of-control ARL is not shown. This approach has the following disadvantage: A control chart has to be constructed for each quality characteristic.

Chiu and Kuo present in their paper the Multivariate Poisson Chart (MP) Chart. This chart is developed by an exact probability method based on the sum of defects or non conformities for each quality characteristic. Their goal is to monitor the correlated multivariate Poisson count data.

When more than one quality characteristic is of interest the MP control chart is needed.

Let X_j the number of non conformities with respect to quality characteristic j, $j=1,2,\ldots,p$. Data X=(X_1, $X_2,\ldots,X_p$) follows a jointly p-variate Poisson distribution. Each X_j follows a Poisson distribution marginally with mean (λ_j), the covariance between two variables (X_j,X_k) is θ_0 $j \neq k$ and D is the sum of all X_j:

$$D = \sum_{j=1}^{p} x_j \text{ , } j=1,2,\ldots,p$$

If $p=2$ the probability function of D is:

$$P(D=d)=\exp[-(\lambda_1+\lambda_2-\theta_0)]\sum_{i=0}^{d/2}\frac{(\lambda_1+\lambda_2-2\theta_0)^{d-2i}\,\theta_0^{i}}{(d-2i)!\,i!},\ d=1,2,\ldots,\infty$$

If $p=3$ the probability function of D is:

$$P(D=d)=\exp[-(\lambda_1+\lambda_2+\lambda_3-2\theta_0)]\sum_{i=0}^{d/3}\frac{(\lambda_1+\lambda_2+\lambda_3-3\theta_0)^{d-3i}\,\theta_0^{i}}{(d-3i)!\,i!},\ d=1,2,\ldots,\infty$$

$$P(D=d)=\exp\{-[\sum_{j=1}^{p}\lambda_j-(p-1)\theta_0]\}\sum_{i=0}^{d/p}\frac{(\sum_{j=1}^{p}\lambda_j-p\theta_0)^{d-pi}\,\theta_0^{i}}{(d-pi)!\,i!},$$

$d=1,2,\ldots,\infty$

The control limits then can be determined by taking the upper and lower α/2 percentage points of the exact distribution in the above equation. Thus, the upper control limit (UCL) and lower control limit (LCL) must be found to satisfy the following properties:

$$P(D>UCL)=\sum_{d=UCL}^{\infty}\exp\{-[\sum_{j=1}^{p}\lambda_j-(p-1)\theta_0]\}\sum_{i=0}^{d/p}\frac{(\sum_{j=1}^{p}\lambda_j-p\theta_0)^{d-pi}\,\theta_0^{i}}{(d-pi)!\,i!}\le\frac{a}{2}$$

$$=1-\sum_{d=0}^{UCL-1}\exp\{-[\sum_{j=1}^{p}\lambda_j-(p-1)\theta_0]\}\sum_{i=0}^{d/p}\frac{(\sum_{j=1}^{p}\lambda_j-p\theta_0)^{d-pi}\,\theta_0^{i}}{(d-pi)!\,i!}\le\frac{a}{2}$$

And

$$P(D<LCL)=\sum_{d=0}^{LCL}\exp\{-[\sum_{j=1}^{p}\lambda_j-(p-1)\theta_0]\}\sum_{i=0}^{d/p}\frac{(\sum_{j=1}^{p}\lambda_j-p\theta_0)^{d-pi}\,\theta_0^{i}}{(d-pi)!\,i!}\le\frac{a}{2}$$

The multivariate Poisson with small mean values for each variable X_j leads to a small or zero value of statistic D. This would cause the probability $P(D=0)$ larger than $a/2$ most of the time. If $P(D=0)>\alpha/2$, LCL is set to 0. If false alarm rate is α UCL is:

$$P(D>UCL)=\sum_{d=UCL}^{\infty}\exp\{-[\sum_{j=1}^{p}\lambda_j-(p-1)\theta_0]\}\sum_{i=0}^{d/p}\frac{(\sum_{j=1}^{p}\lambda_j-p\theta_0)^{d-pi}\,\theta_0^{i}}{(d-pi)!\,i!}<a$$

$$=1-\sum_{d=0}^{UCL-1}\exp\{-[\sum_{j=1}^{p}\lambda_j-(p-1)\theta_0]\}\sum_{i=0}^{d/p}\frac{(\sum_{j=1}^{p}\lambda_j-p\theta_0)^{d-pi}\,\theta_0^{i}}{(d-pi)!\,i!}<a$$

For λ_j and θ_0 their values can be estimated from the observed sample data. When UCL is determined, we can plot D on the MP chart and monitor the process.

Xie et al. (2002) pointed out that the normal distribution can be used to approximate Poisson distribution when the mean value is greater than five. Similar results can be found in our simulation study. When $D = \sum_{j=1}^{p} \lambda_j$ is greater than five, normal approximation is good. By normal approximation, traditional Shewhart type control limits can be obtained. The Shewhart-type control limits and centerline (CL) are below:

$$UCL = \sum_{j=1}^{p} \lambda_j + 3\left(\sum_{j=1}^{p} \lambda_j + 2\sum \rho_{jk} \sqrt{\lambda_j \lambda_k}\right)^{1/2}$$

$$CL = \sum_{j=1}^{p} \lambda_j$$

$$LCL = \sum_{j=1}^{p} \lambda_j - 3\left(\sum_{j=1}^{p} \lambda_j + 2\sum \rho_{jk} \sqrt{\lambda_j \lambda_k}\right)^{1/2}$$

The performance of the MP control chart is investigated with the help of ARL_1 and ARL_0. The results of this comparison are that MP chart has ARL_0 closer to the nominal ARL than the Shewhart type chart. Therefore, the MP chart is more sensitive to large correlation coefficients if the process is out-of-control.

5.3 Generalized Linear Model based control charts for multiple count data

A procedure for monitoring multiple discrete counts proposed by Skinner, Montgomery and Runger is based on the likelihood ratio statistic for Poisson counts when input variables are measurable. This is actually the deviance residual resulting from a GLM. This statistic is more effective than a C chart on the raw counts.

Unlike model based procedures proposed for monitoring multivariate means the procedure examined by Skinner, Montgomery and Runger does not require that the variables of interest are normally distributed. However, it retains the benefits of other model based procedures which are that:

- . the control statistics have relatively simple interpretations;
- the control statistics are based on residuals from a model and are not usually correlated over time; and
- the methods are easy to perform, requiring only a least squares model-fitting program.

Ordinary least squares regression analysis is the basis for model based charts. The effectiveness of these monitoring schemes is limited when dealing with non normal data. Because non-normal data are common in many industrial applications, a GLM can be used to unify the fields of linear and nonlinear regression and to include the ability to model responses from many different distributions. A specific GLM for normal data is the same model that is obtained with OLS, yet GLM is not limited to use with normally distributed data.

Generalized Linear Models are used to model a response variable that is a member of the exponential family, which includes the normal, binomial, exponential, Poisson, and gamma distributions. The model parameters link the response distribution to the regressor variables through the linear function $x_i'\beta$, a linear predictor of μ_i=$\underline{E(y_i)}$. The link function is given by

$$x_i'\beta=s(\mu_i).$$

The GLM predicted value is the mean given by the inverse of the link function of the linear predictor. For Poisson data, common link functions are the log link, the square root link, and the inverse link; conversely, common inverse link functions are exponential, square, and inverse.

Skinner, Montgomery and Runger propose a monitoring scheme based on the deviance residuals. Under H_0, the deviance residuals for the GLM are independent and asymptotically normally distributed. They are ideal for plotting on p Shewhart charts for individuals. Their scheme uses an empirical GLM based on input variables to predict the output variables, or counts. The steps are:

Step 1. Fit the GLM to each of the *p* Poisson counts, choosing the link function to obtain the best fit.
Step 2. Obtain predictions for future values.
Step 3. Calculate deviance residuals calculated from above equation.
Step 4. Monitor these deviance residuals on p Shewhart charts for individuals. A signal by any one of the p Shewhart charts indicates that the variable it monitors is out of control.

Skinner, Montgomery and Runger tested 3 different kinds of shifts: a univariate case, a bivariate case with equal means and a bivariate case with unequal means and all shifts could be detected by the deviance residual. The deviance residual from the GLM outperforms univariate and bivariate C charts.

CHAPTER 6.

CONCLUSIONS

Modern Businesses main aim is to improve quality of the products and services they provide. However, almost always, variability exists, making each product different from the other. A control chart is a tool that contributes in the monitoring of the variability of a process, so as to detect the occurrence of assignable causes of process shifts as quickly as possible. If a shift and its cause are detected, corrective actions can take place before many non-conforming items are produced.

In every process an amount of natural variability cannot be avoided. A process is said to be ‘in control’ if only chance causes of variation exist. However, if other causes of variability that are not part of the chance exist, the process that is operating in the presence of assignable causes is ‘out of control’.

In order for an item to be nonconforming it has to contain at least one nonconformity. Of course, an item with only a few insignificant nonconformities can still remain conforming. Often the number of nonconformities is as or even more important than its characterization as conforming or nonconforming.

The occurrence of nonconformities in samples of constant size is well modelled by the Poisson distribution. This implies that the number of opportunities for the occurrence of nonconformities has to be infinity and that the probability of occurrence of nonconformities should be a constant (λ).

This thesis is a review of papers that present Control Charts for Poisson Data. Khoo compares the C Chart for Nonconformities to the Poisson Moving Average chart and it seems that the Moving Average outperforms the traditional C Chart in terms of in-control and out-of-control ARLs.

The modified U Control Chart of Rudisill, Litteraland Walter is very effective when other sources of variability apart from the Poisson exist.

In automatic high yield manufacturing and continuous production processes a process has usually an extensive count of zeros, more than expected under its underlying Poisson distribution. Therefore, a model was developed as a zero defect process subject to random shocks. The random shock occurs with probability p and upon the occurrence of a random shock, non-conformities can be found. This number of nonconformities follows the Poisson distribution. This distribution is called the Zero Inflated Poisson Distribution.

An alternative of creating a control chart for the ZIP distribution is to construct the upper control limit of the ZIP model based on the Jeffreys prior interval. This would provide good coverage probability for λ.

A generalized zero inflated Poisson distribution, an extension of the ZIP model is also presented. The GZIP model is a particular form of the Poisson distribution.

In addition, several EWMA as well as CUSUM control charts are being examined. The Double EWMA chart requires a second Exponentially Weighted Moving Average.

Sheu and Lin (2008) presented, the Generally Weighted Moving Average Control Chart. The purpose of developing this chart would be to monitor Poisson observations and detect persistent, small process shifts. PEWMA chart and c-chart are special cases of the PGWMA generalized charting model. In addition other PGWMA control charts are being presented, such as ACLPGW, a PGWMA chart with asymptotic control limits, VCLPGW, a PGWMA chart with time-varying control limits, FVCLPGW, a FIR (Fast Initial Response) PGWMA chart with time-varying control limits, FADJPGW, another FIR (Fast Initial Response) PGWMA chart with time varying control limits and PDGWMA a Poisson Double GWMA control chart

Moreover, the conditional and marginal performance of a Poisson CUSUM chart with an unknown parameter is examined.

Chen Randolph and Liou’s paper, presenting a CUSUM control chart based on the Poisson distribution compounded by a Geometric distribution is also introduced.

Finally, in the last chapter, a model-based control chart for a multivariate Poisson with measurable inputs proposed by Skinner et al. (2003) is presented. The independence between output variables, indicates that the deviance residuals result in a generalized linear model that is used in the monitoring scheme for each variable. For correlated multivariate Poisson count data Chiu and Kuo introduced the multivariate Poisson chart.

REFERENCES

Abbasi B., Taghi S. (2007). Monitoring high-yields processes with defects count in nonconforming items by artificial neural network, *Applied Mathematics and Computation*, 188, 262–270

Borror C., Champ Ch., Rigdon S. (1998). Poisson EWMA Control Charts, *Journal of Quality Technology*, 30 no4, 352-61

Chen Ch., Randolph P., Liou T. (2005). Using CUSUM Control Schemes for Monitoring Quality Levels in Compound Poisson Production Environment: The Geometric Poisson Process S, *Quality Engineering*, 17, 207–217

Chen N., Zhou S., Chang T., Huang H. (2008). Attribute Control Charts Using Generalized Zero-inflated Poisson Distribution, *Quality and Reliability Engineering. Int.*, 24, 793–806

Chiu J., Kuo T. (2008). Attribute Control Chart for Multivariate Poisson Distribution, *Communications in Statistics—Theory and Methods*, 37, 146–158

Chiu W., Sheu S. (2008). Fast Initial Response Features for Poisson GWMA Control Charts, *Communications in Statistics—Simulation and Computation*, 37, 1422–1439

H¨ohlea M., Paulc M. (2008). Count data regression charts for the monitoring of surveillance time series, *Computational Statistics and Data Analysis*, 52, 4357–4368

Khoo M. (2004). Poisson Moving Average Versus c Chart for Nonconformities, *Quality Engineering*, 16, No. 4, 525–534

Montgomery D. (2001). *Introduction to Statistical Quality Control*. John Wiley & Sons, New York

Perry M., Pignatiello J., Simpson J. (2007). Change point estimation for monotonically changing Poisson rates in SPC, *International Journal of Production Research*, 45, No. 8, 1791–1813

Perry M., Pignatiello J., Simpson J. (2006). Estimating the Change Point of a Poisson Rate Parameter with a Linear Trend Disturbance, *Quality and Reliability Engineering International*, 22, 371–384

Rudisill F., Litteral L., Walter D. (2004). Modified U Charts for Monitoring and Controlling Poisson Attribute Processes, *Quality Engineering*, 16, No. 4, 637–642

Sheu S, Chiu W. (2007), Poisson GWMA Control Chart, *Communications in Statistics—Simulation and Computation*, 36, 1099–1114

Sim C., Lim H. (2008), Attribute Charts for Zero-Inflated Processes, *Communications in Statistics—Simulation and Computation*, 37: 1440–1452

Skinner K., Montgomery D., Runger G. (2003). Process monitoring for multiple count data using generalized linear model-based control charts, *International Journal of Production Research*, 41, No. 6, 1167–1180

Testik M. (2007). Conditional and marginal performance of the Poisson CUSUM control chart with parameter estimation, *International Journal of Production Research*, 45, No. 23, 5621–5638

Weiss Ch. (2007). Controlling Correlated Processes of Poisson Counts, *Quality and Reliability Engineering International*, 23: 741–754

Xie M., He B., Goh T. (2001). Zero-inflated Poisson model in statistical process control, *Computational Statistics & Data Analysis*, 38, 191–201

Zhang L., Govindaraju K., Lai C., Bebbington M. (2003), Poisson DEWMA Control Chart, *Communications in Statistics Simulation and Computation*, 32, No. 4, 1265–1283