

Control Chart Based on Transition Probability Approach

K. Thaga¹ and R. Sivasamy²

Abstract

A Transition Probability approach is developed for constructing control charts to monitor attribute processes when sample data sets are collected in linguistic forms. Resulting performance of Markov chain (MC) theory based control chart called transition probability control chart (TPCC) is compared with that of the membership approach based control chart called fuzzy control chart (FCC). A numerical example is given to illustrate the application of the proposed control charts to check if the FCC performs better than the TPCC in monitoring the quality characteristics of a production process.

Keywords: Product Quality; Linguistic Terms; Fuzzy Control Chart (FCC); Markov Dependent Samples; Transition Probability Control Charts (TPCC).

¹ University of Botswana. E-mail: thagak@mopipi.ub.bw

² University of Botswana. E-mail: ramasamy@mopipi.ub.bw

1 Introduction

The primary objective of engineers and manufacturers who monitor complex production lines is to produce quality items while the production process is operating under stable conditions. Such a goal is achieved by installing appropriate statistical packages in a few modern on-line computers. Any procedure of monitoring that uses statistical techniques to control qualitative or quantitative characteristics of a product is called statistical process control (SPC) method. For example, control charts are frequently used as monitoring tools for monitoring quality characteristics of manufactured products over a period of time to ensure that quality goods are produced.

In 1924, Walter Shewhart, introduced the control chart which plots the observed data relating to one or more characteristics (variables or attributes) of interest of a product produced by a process on a chart showing the target values, an upper control limit (UCL) and a lower control limit (LCL) to constitute a crisp picture. A simple inspection over the chart enables us to identify whether or not the process is in control. If a few data points fall outside the UCL or LCL, we say that a possible out of control condition due to assignable causes operates within the process. “Out of control” conditions mean that the process is producing products that are not close to the target value.

Crisp Set: All possible realizations of a quantitative variable like ‘weight or length of an item’ is a numerical value which we call a crisp number and the collection of such crisp values is called a crisp set. Hence a crisp set has clearly defined boundaries. Each set ‘A’ of the classical set theory has well defined clear boundaries and the membership of elements in the set A can be assessed by the binary terms 0 and 1 and hence a crisp number $CN(a)=1$ if $a \in A$ and $CN(a)=0$ if $a \notin A$ are the two true values assigned by the classic propositional logic theory.

Fuzzy set Theory: Fuzzy sets generalize the classical sets by permitting the gradual assessment of the membership of elements in the real interval $[0, 1]$ through a rule defined by a membership function. Using “Fuzzy theory”, one can measure

linguistic terms with reasoning that is approximate rather than accurate. For example, most attribute characteristics of a product that are observed through quality levels or performance types of a product such as good, and bad are not crisp values and hence they are categorized as high standard, medium standard or poor standard are called linguistic terms of the linguistic variable quality level or performance level. Though fuzzy logic has been successfully applied to many fields, it is also considered as a controversial tool among most statisticians who prefer crisp logic. Due to non-clarity or vagueness associated with those terms of the linguistic variable, mathematical rules cannot assign a unique crisp value to any of the linguistic terms but can be transformed into crisp values using the fuzzy set theory introduced by Zadeh (1965).

Researchers have suggested that using the binary classification to measure the degree of satisfaction on the quality levels of a product into confirming (by 1) and nonconforming (by 0) states as used in the p-chart is not appropriate since there might be a number of intermediate levels and hence the degree of satisfaction can vary from 0 to 1. For example attribute characteristics of a product can be classified into 'perfect, good, fair and poor' instead of classifying each product either as good (confirming) and poor (nonconforming). Out of three human judgements made on a product, two judgements may classify its quality level as perfect and the third judgment could classify the same product as either fair or poor since there are no clearly defined boundaries between perfect and good categories or between fair and poor categories. Hence it is not possible to assign a unique crisp value to each of those quality levels 'perfect, good, fair and poor' of any product with certainty by human judgements or rules due to the presence of undefined boundaries or vagueness or imprecision with each of the quality levels but these can be well handled by fuzzy set theory.

Fuzzy Set F: A fuzzy set F defined in a collection X of base objects x contained in a Universe of discourse U say, is a set of ordered pairs

$$F = \{ (x, \mu_F(x) \in M) : x \in X \} \quad (1)$$

$\mu_F(x)$ is called the membership function which maps each member x of the base variable X into a bounded membership space M of non-negative real numbers, whose supreme(sup) value is finite. Also, F is called a normalized fuzzy set if

$$0 \leq \mu_F(x) \leq 1 \text{ and Mode value of } F=x_0 \text{ if } \sup \mu_F(x_0) = 1 \quad (2)$$

A Convex Fuzzy Set F: A Fuzzy set $F = \{ (x, \mu_F(x) \in M) : x \in X \}$ is convex if

$$\mu_F(\theta x_1 + (1-\theta)x_2) \geq \min\{\mu_F(x_1), \mu_F(x_2)\}, \quad x_1, x_2 \in X, \text{ and } \theta \in [0,1] \quad (3)$$

A triangular fuzzy set F: A Fuzzy set $F = \{ (x, \mu_F(x) \in M) : x \in X \}$ is called triangular type if it can also be assigned a TFN(triangular fuzzy number) $\tilde{F} = (a, b, c)$, 'b' being the mode of F . If $b-a=c-b$ or $b=(a+c)/2$, it is said to be symmetrical; otherwise asymmetrical. Note that each fuzzy number is also a fuzzy set with $\mu_F(a) = 0 = \mu_F(c)$ and $\mu_F(b) = 1$.

For constructing control charts firstly a methodology to determine process characteristics is to be developed and secondly a suitable approach is needed to determine the various parameters of the associated control chart. Using the concepts of the Fuzzy set theory, various types of fuzzy control charts have been proposed by Raz and Wang (1990), Kanagawa et al. (1993), Taleb and Limam (2002), Gulbay et al. (2004), Cheng (2005), Hryniewicz(2007), Faraz et al. (2009), Demirli and Vijayakumar(2010), and Shu and Wu (2011) to deal with uncertainties associated with categorical data on linguistic terms. Fuzzy control charts developed by Feili and Fekraty (2010), Wang and Raz (1990), Gulbay et al. (2004) and Sorooshian (2013) have some important advantages compared to Shewhart control charts in handling such uncertainties.

A basic requirement for the construction of attribute control charts in each of these articles is that an adequate sample size should be selected to include at least one item in each category of the linguistic variable under consideration such that the normality assumption is not violated. Nevertheless if the output rate of the production process is small then large sample size selection becomes impossible or becomes time consuming and costly. To overcome this drawback this article

develops a transition probability approach for constructing fuzzy control charts to monitor attribute levels of processes when data on small sample sizes are collected in linguistic terms (LTs) such as LT_1 =High Standard, LT_2 =Medium Standard, LT_3 =Low Standard, and LT_4 =Poor Standard. This paper provides a few illustrations by creating hypothetical sample data sets on these LT_1 , LT_2 , LT_3 and LT_4 .

Further constructing ‘Fuzzy Control Charts’ similar to that of Shewhart control charts, a ‘normalized Fuzzy Set’ to each linguistic term LT_j for $j=1, 2, 3,$ and 4 is assigned as in Raz and Wang, based on the membership functions (4) that are standardized in $[0,1]$ for the evaluation of product quality in terms of a base variable ‘x’:

$$\mu_{LT_1} = \begin{cases} 0 & \text{for } x < 0 \\ -2x+1 & \text{for } 0 \leq x \leq 0.5 \\ 0 & \text{for } x \geq 0.5 \end{cases} \quad \mu_{LT_4} = \begin{cases} 0 & \text{for } x < 0.5 \\ 2x-1 & \text{for } 0.5 \leq x \leq 1 \\ 0 & \text{for } x \geq 1 \end{cases}$$

$$\mu_{LT_2} = \begin{cases} 0 & \text{for } x < 0 \\ 4x & \text{for } 0 \leq x \leq 0.25 \\ -2x + \frac{3}{2} & \text{for } 0.25 \leq x \leq 0.75 \\ 0 & \text{for } x \geq 0.75 \end{cases} \quad \mu_{LT_3} = \begin{cases} 0 & \text{for } x < 0.25 \\ 2x & \text{for } 0.25 \leq x \leq 0.5 \\ -2x + 2 & \text{for } 0.5 \leq x \leq 1 \\ 0 & \text{for } x \geq 1 \end{cases} \quad (4)$$

If $\mu_F(x)$ is 1 for zero value of x then the term that is represented by F corresponds to the best quality and if $\mu_F(x)$ is 1 for $x=1$, then the term that is represented by F corresponds to the poor quality.

The fuzzy set F_{LT_j} that identifies the term LT_j for $j=1, 2, 3,$ and 4 through μ_{LT_j} of (4) is called a triangular fuzzy set which can also be assigned a fuzzy number $\tilde{F}_{LT_j} = (a_j, b_j, c_j)$, ‘b’ being the mode of F_{LT_j} : $\tilde{F}_{LT_1} = (0, 0, 0.5)$, $\tilde{F}_{LT_2} = (0, 0.25, 0.75)$, $\tilde{F}_{LT_3} = (0.25, 0.5, 1)$ and $\tilde{F}_{LT_4} = (0.5, 1, 1)$ and thus each of them can be graphically represented by a corresponding triangular shape as shown in Figure 1.

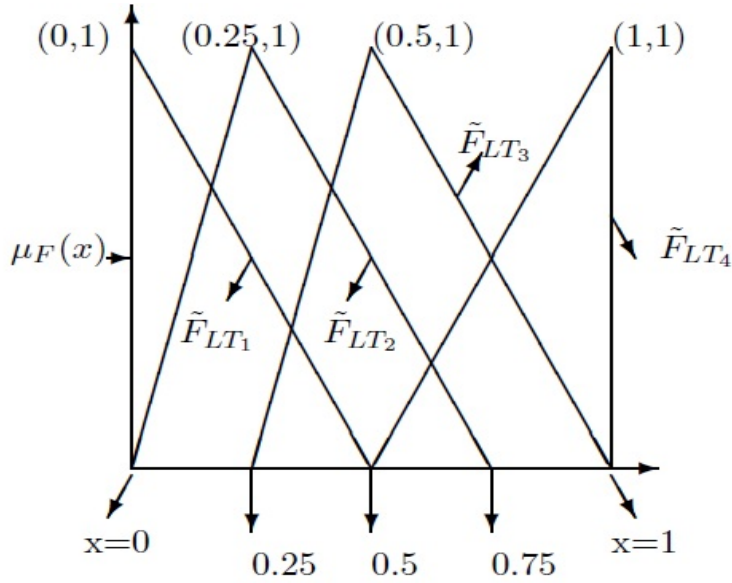


Figure 1: Triangular graph of fuzzy sets F_{LT_j} for $j=1, 2, 3$, and 4

Fuzzy Averages: It is then necessary to compute averages which are called representative values as crisp values for each of the above fuzzy sets $\tilde{F}_{LT_j} = (a_j, b_j, c_j)$. For such a conversion to each of those fuzzy sets associated with the linguistic terms into a scalar, to act as a *representative value*, here two ways which are similar in descriptive statistics are presented, assuming the membership functions are nonlinear:

1. *Fuzzy mode* (f_{mod}): The fuzzy mode of a fuzzy set F with $\tilde{F} = (a, b, c)$ and membership function $\mu_F(x)$ is b which also satisfies $\mu(b) = 1$.
2. *Fuzzy median* (f_{med}): The fuzzy median (f_{med}) of a fuzzy set ($F, \mu_F(x)$) with $\tilde{F} = (a, b, c)$ is the point which partitions the curve under the membership function $\mu_F(x)$ of F into equal regions which leads to the following value,

$$f_{med} = \begin{cases} c - \sqrt{\frac{(c-a)(c-b)}{2}} & \text{for } b < \frac{c+a}{2} \\ a + \sqrt{\frac{(c-a)(b-a)}{2}} & \text{for } b > \frac{c+a}{2} \end{cases} \quad (5)$$

Thus the representative value r_j (either f_{mod} or f_{med}) using the membership function of each fuzzy subset representing a linguistic term LT_j for $j=1, 2, 3$ and 4 of (4) are computed and reported in Table 1:

Table 1: Representative Values r_j of membership functions of $\mu_{LT_j}(x)$ for $j=1,2,3,$ and 4

Categories 'j'	Fuzzy Mode (r_j)	Fuzzy Median(r_j)
LT_1 or 1	0.0	0.146
LT_2 or 2	0.25	0.317
LT_3 or 3	0.5	0.567
LT_4 or 4	1.0	0.854

Sampling Distribution: Let 'm' be the number of samples and X_{kj} be the number of categories of LT_j found in the k^{th} sample for $j=1, 2, 3,$ and 4 so that the k^{th} sample size is $\sum_{j=1}^4 X_{kj} = n_k$. Then r_j is representing each LT_j ($j=1, 2, 3, 4$) in the k^{th} sample for $k=1, 2, \dots, m$, which is not dependent on 'k'.

The rest of the paper is organized as follows: Section 2 gives a brief overview on the Markov dependent samples and their transition probability matrices, probability distribution for the representative values of all categories represented by fuzzy sets and with TPCC. Section 3 outlines the way of constructing the generalized p-chart in terms of large sample sizes generated by the MC theory. Section 4 addresses basic issues of the fuzzy mode, mean deviation of the process under study as fuzzy numbers and then the construction of control chart for the

expected number of defects in the long run estimated by the MC methods. The paper is concluded outlining the uses of MC approach in section 5.

2. Markov dependent samples and their transition probability matrices

Using the observed attribute data on linguistic terms of a product, one can verify if each item conforms to the standards and its degrees of satisfaction is high or low. As those linguistic terms of any production process tend to be time dependent if they are observed over a period of time, it is necessary to employ an appropriate tool like 'Markov dependent samples' due to Sivasamy and Jayanthi (2006) to take into account the effects of time dependent relationships among the samples sizes. Hence, this section formulates an alternative methodology based on transition probability matrices of Markov Chains (MCs) for constructing control charts using fuzzy representatives of linguistic terms.

2.1 Selection of Markov Dependent Samples

Let the size n_k of the k^{th} sample be a finite integer for $k=0,1,2,\dots,m$. Assume that the categories of these n_k sampling units observed over a period of time form an ergodic Markov chain on the state space $\mathbb{S}=\{1=\text{High Standard}(LT_1), 2=\text{Medium Standard}(LT_2), 3=\text{Low Standard}(LT_3), 4=\text{Poor Standard}(LT_4)\}$ with unit step transition probability matrix (TPM) $\mathbf{P}_k=(p_{ij})$, where p_{ij} denotes the unit step conditional transition probability of producing the j^{th} category starting from the i^{th} category in the k^{th} sample. Let p_{ij}^r denote the r -step conditional transition probability of moving to the j^{th} category starting from the i^{th} category in the k^{th} sample and π_{kj} be the stationary probability of j^{th} category (for $j = 1, 2, \dots, t$) of the

k^{th} sample in the long run which can be obtained from $\pi_{kj} = \lim_{r \rightarrow \infty} {}_k P_{ij}^r$. Let $\boldsymbol{\pi}_k = (\pi_{k1}, \pi_{k2}, \pi_{k3}, \pi_{k4})$ be a row vector and $\mathbf{e}' = (1,1,1,1)$ be column vector of unit elements. If $\pi_{kj} = \lim_{r \rightarrow \infty} {}_k P_{ij}^r > 0$, then π_{kj} values can be computed by solving the following system of equations:

$$\boldsymbol{\pi}_k \mathbf{P}_k = \boldsymbol{\pi}_k \text{ and } \boldsymbol{\pi}_k \mathbf{e}' = 1 \tag{6}$$

Let X_{kj} = number of items of the j^{th} category in the Markov dependent sample selected at the k^{th} period with size $\sum_{j=1}^4 X_{kj} = n_k$, and let $\mathbf{x}_k = (X_{k1}, X_{k2}, X_{k3}, X_{k4})$ be a vector of size four.

2.2 Markov Dependent Sample sizes

Suppose that a production process is in control in the period ‘ $k=0$ ’ and yields the following results over the space above state space \mathbb{S} :

$$\begin{aligned} & (k=0, \\ & n_0=18): 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow 4 \rightarrow 3 \end{aligned} \tag{7}$$

The sample ($k=0, n_0=18$) so selected by (7) is called the reference sample in which only 2 items are of poor standard category out of a total 18. A simple inspection over the observed sequence of states of the MC in the sample (7) enables us to obtain the TPM \mathbf{P}_0 , sample $\mathbf{x}_0 = (X_{01}, X_{02}, X_{03}, X_{04})$, and its stationary distribution $\boldsymbol{\pi}_0$, as below:

$$\mathbf{P}_0 = \begin{pmatrix} 5/7 & 2/7 & 0 & 0 \\ 2/5 & 0 & 2/5 & 1/5 \\ 0 & 2/3 & 1/3 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix},$$

$$\mathbf{x}_0 = (7,5,4,2) \text{ and } \boldsymbol{\pi}_0 = (0.3783784, 0.2702703, 0.2432432, 0.1081081) \tag{8}$$

The following are the additional summary statistics obtained from $m=10$ more samples of periods as it was done to compute (8) relating to the sequence of Markov dependent sample of size n_k selected for $k=1, 2, \dots, 10$, together with n_k , x_k and π_k :

($k=1, n_1=18$): $2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow 4 \rightarrow 3$

$x_1=(6,6,4,2)$ and $\pi_1=(0.3243243,0.3243243,0.2432432,0.1081081)$

($k=2, n_2=12$): $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 2$

$x_2=(2,5,3,2)$ and $\pi_2=(0.1086957,0.4347826,0.2608696,0.1956522)$

($k=3, n_3=12$): $3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 4 \rightarrow 1 \rightarrow 3 \rightarrow 2$

$x_3=(3,4,3,2)$ and $\pi_3=(0.3,0.3375,0.1875,0.175)$

($k=4, n_4=14$): $1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 4 \rightarrow 3 \rightarrow 1$

$x_4=(7,3,3,1)$ and $\pi_4=(0.4615385,0.2307692,0.2307692,0.07692308)$

($k=5, n_5=19$): $2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow 4 \rightarrow 3 \rightarrow 1$

$x_5=(5,5,6,3)$ and $\pi_5=(0.2934363,0.2316602,0.3243243,0.1505792)$

($k=6, n_6=7$): $2 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 1 \rightarrow 2$

$x_6=(1,3,1,1)$ and $\pi_6=(0.1666667,0.5,0.1666667,0.1666667)$

($k=7, n_7=12$): $4 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 3 \rightarrow 1$

$x_7=(3,2,4,3)$ and $\pi_7=(0.3478261,0.173913,0.3478261,0.1304348)$

($k=8, n_8=13$): $1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 4 \rightarrow 3$

$x_8=(6,3,3,1)$ and $\pi_8=(0.3333333,0.25,0.2777778,0.1388889)$

($k=9, n_9=8$): $2 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 4$

$x_9=(1,4,1,1)$ and $\pi_9=(0.125,0.25,0.125,0.5)$

($k=10, n_{10}=15$): $2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 4 \rightarrow 3$ (9)

$x_{10}=(6,4,3,1)$ and $\pi_{10}=(0.3333333,0.25,0.2777778,0.1388889)$

The probability distribution for the representative values f_{mod} of all categories of the k^{th} sample that were reported in Table 1 is shown in Table 2:

Table 2: Probability distribution of r_j of linguistic terms for $j=1, 2, 3$ and 4

Sample k:	$0=r_1$	$0.25=r_2$	$0.5=r_3$	$1=r_4$
Probability :	π_{k1}	π_{k2}	π_{k3}	π_{k4}

The values of following statistics may now be computed easily from Table 3:

$$M_k = \text{mean of the } k^{\text{th}} \text{ sample} = \sum_{j=1}^4 \pi_{kj} r_j \text{ for } k=1, 2, \dots, m \text{ and } CL = \frac{1}{m} \sum_{k=1}^m M_k \quad (11)$$

The standard deviation and the mean of standard deviations (MSD) of these m samples is

$$SD_k = \sqrt{\sum_{j=1}^4 \pi_{kj} (r_j - M)^2} \text{ for } k=1, 2, \dots, (m=10) \text{ and } MSD = \frac{1}{10} \sum_{k=1}^{m=10} SD_k \quad (12)$$

From (11) and (12) it is noticed that sample means M_k of representative values $\{r_j\}$, the SD_k values and thus MSD should lie within the range $[0,1]$. The centreline (CL) of the control chart is now defined as the grand average of means $\{M_k: k=1, 2, \dots, 10\}$.

2.3 Transition Probability Control Chart (TPCC)

Since $\lim_{k \rightarrow \infty} \frac{X_{kj}}{n_k} = \pi_{kj}$ with the passage of time in the long run, assuming the sample sizes 120, 132, 143, 125, 110, 95, 142, 154, 100, 150 as N_k for $k= 1, 2, \dots, 10$ and the sample size for the reference sample i.e. k_0 as 125 at random.

Table 3: $E(X_{kj})$ for a given N_k ; $k=1$ to 10 and $j=1$ to 4.

k	$E(X_{k1})$	$E(X_{k2})$	$E(X_{k3})$	$E(X_{k4})$	N_k
1	39	39	29	13	120
2	14	57	35	26	132
3	43	48	27	25	143
4	57	29	29	10	125
5	32	26	35	17	110
6	16	47	16	16	95
7	49	25	49	19	142
8	51	39	43	21	154
9	13	50	12	25	100
10	73	35	31	11	150
k=0 (Reference)	47	34	30	14	125

It is now remarked that those randomly fixed values of N_k are sufficiently large to attain the stationary state of the MC under study corresponding to the small n_k , one can calculate the expected number $E(X_{kj})$ approximately for the above samples on LT_j *i.e.*

$$E(X_{kj}) = \pi_{kj} N_k \text{ for } j=1, 2, 3, \text{ and } 4 \quad (13)$$

These values so calculated using (13) have been reported in Table 3 for further discussions.

To apply the standard formulae of variables control charts based on the normal distribution assumption for a sample of size N_k for $k=0, 1, 2, \dots, 10$, A_{3k} and C_{4k} can be found in the table of co-efficient for control charts of Montgomery:

$$A_{3k} = \frac{3}{C_{4k} \sqrt{N_k}} \text{ and } C_{4k} = \frac{4(N_k - 1)}{4N_k - 1} \quad (14)$$

Based on the above stationary probability distributions $\{\pi_{kj} : j=1,2,3 \text{ and } 4\}$ obtained for $k=1, 2, \dots, (m=10)$, and the f_{mod} values $\{r_j : j=1,2,3 \text{ and } 4\}$ of fuzzy subsets reported in Table 3, the sample data points $\{M_k\}$ and $CL(=0.345)$ value of (11) and the $UCL_k = \text{Max}\{0, (CL - A_{3k} \text{ MSD})\}$ and $LCL_k = \text{Min}\{0, (CL + A_{3k} \text{ MSD})\}$ using (12) and (13) have been computed using the above Markov dependent samples of the preceding section. The numerical results of these have been displayed in the form a control chart called the ‘Transition Probability Control Chart’ in Figure 1.

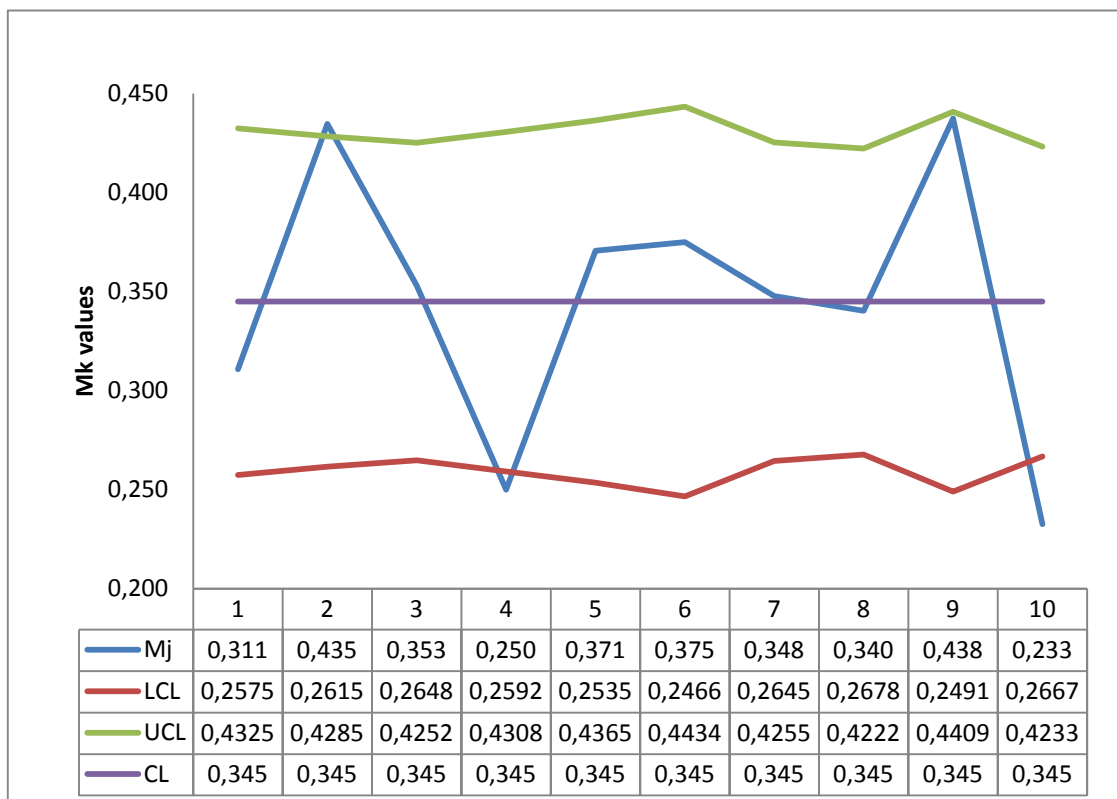


Figure 2: Transition Probability Control Chart

The data points M_k corresponding to sample numbers $k=2, 4$ and $k=10$ fall outside the UCL/LCL due to some assignable causes of variation and sample $k=$ falls on the UCL border.

3 Generalized ‘p’ chart

Different procedures are proposed to monitor two or more characters simultaneously, when products are classified into mutually exclusive categories. The one-sided monitoring of quality proportions is designed to detect only an increase in all but one quality proportions. An appropriate statistical procedure when specific values of process proportions are not known that is a test of homogeneity of proportions between the base period (0) and each monitoring period (i) is defined as follows for $i=1, 2, \dots, m$:

$$Z_i^2 = \sum_{k \in \{0,i\}} \sum_{j=1}^t \frac{N_k \left\{ \frac{E(X_{kj})}{N_k} - \frac{E(X_{ij}) + E(X_{0j})}{N_i + N_0} \right\}^2}{\frac{E(X_{ij}) + E(X_{0j})}{N_i + N_0}} = N_0 N_i \sum_{j=1}^t \frac{(\pi_{ij} - \pi_{0j})^2}{E(X_{ij}) + E(X_{0j})} \quad (15)$$

where $k \in \{0,i\}$, $E(X_{kj})$ = expected number of items of the j^{th} category in the Markov dependent sample selected at the k^{th} period with size $\sum_{j=1}^4 E(X_{kj}) = N_k$, and π_{kj} is the stationary probability of j^{th} category for $j = 1, 2, 3$ and 4 in the k^{th} sample so that $E(\mathbf{x}_k) = (E(X_{k1}), E(X_{k2}), E(X_{k3}), E(X_{k4}))$ is a vector of size four.

If $n_i \rightarrow \infty$ where n_0/n_i is finite and greater than zero, so that Z_i^2 has a chi-square distribution with three degrees of freedom. Therefore, the control chart for this case also has an upper control limit equal to an appropriate percentile of chi-square distribution.

There is no theoretical rule for sufficient sample size for using chi-square distribution in the above case. Some rules of thumb exist to determine adequate sample size in Cochran (1954). He declared that the twenty per cent of the frequency of each category should be greater than 5, and the expected frequency of each category should be greater than one.

The generalised p control chart is displayed below in Figure 3 with its data points Z_k^2 that are computed using the values of Table 2 and an upper control limit

equal to 3(5% percentile of chi-square distribution=7.815 for 3 three degrees of freedom) where 3 is the scaling factor.

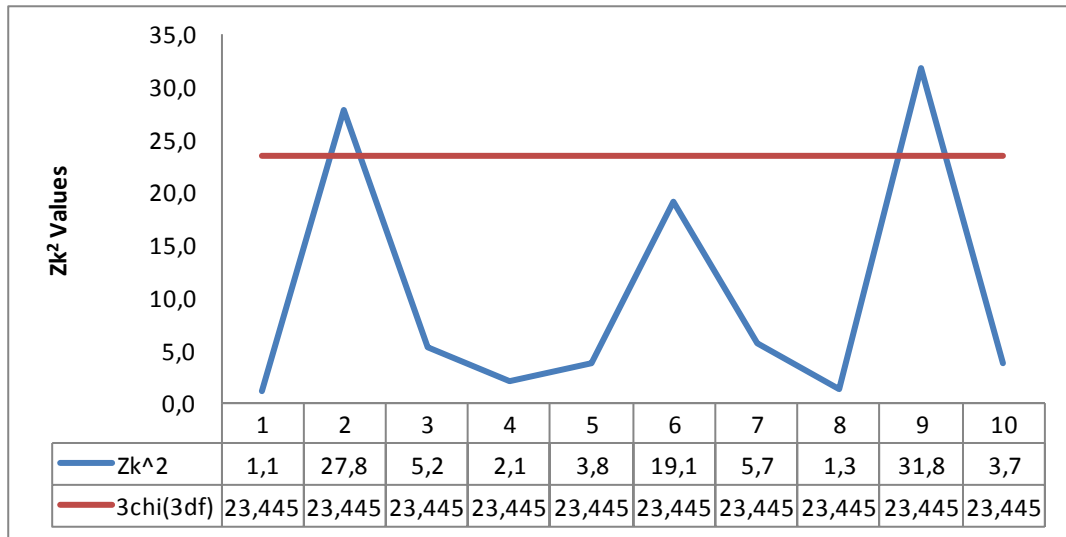


Figure 3: Generalised p Chart

It is noticed from this p chart that data points corresponding to (one-sided) sample-2 and sample-9 fall outside the upper control limit due to some assignable causes of variations. As it is based on a chi-square distribution, there is no specification about lower limit boundary.

4 Fuzzy membership approach based control charts

Let $(F, \mu_F(x))$ be a convex fuzzy set and ' x_{mode} ' be the fuzzy mode so that $\mu_F(x_{mode}) = 1$, if the membership function of F is $\mu_F(x) \in [0,1]$ for $x \in [0, 1]$; the base variable 'x' is being standardized. It is remarked that a triangular type of fuzzy set could also be stated by a triangular fuzzy number (TFN). Let $x_l(\alpha)$ be the inverse function taking the values that fall to the left of x_{mode} and $x_r(\alpha)$ be the inverse function taking the values that fall to the right of x_{mode} for $\alpha \in [0,1]$ when $x_l(\alpha)$ is the

minimum value of the base variable x for which $\mu_F(x) = \alpha$ and $x_r(\alpha)$ is the maximum value of the base variable x for which $\mu_F(x) = \alpha$. Thus $x_l(\alpha) (\leq x_{mode})$ and $x_r(\alpha) (\geq x_{mode})$ are the end points of α -cut while α is the level of membership. The triplet denoted by $F_\alpha = (a, b, c)$ where $a = x_l(\alpha=0)$, $b = x_{mode}$ and $c = x_r(\alpha=0)$ is the TFN of the Fuzzy set when $\alpha=0$.

Deviation of Mean for $(F, \mu_F(x))$: Let the mean deviation of the fuzzy set $(F, \mu_F(x))$ be denoted by $\delta (= \delta_l + \delta_r)$, the sum of $\delta_l =$ deviation of left mean and $\delta_r =$ deviation of right mean values so that:

$$\delta_l = \int_{\alpha=0}^1 [x_{mode} - x_l(\alpha)] d\alpha \text{ and } \delta_r = \int_{\alpha=0}^1 [x_r(\alpha) - x_{mode}] d\alpha$$

Thus the deviation of the fuzzy set $(F, \mu_F(x))$ is a crisp value which can be calculated by the dimension of the base variable 'x'. For $\alpha \neq 0$, the average of the end points of an alpha-cut (α -cut) is defined as the α -level fuzzy midrange.

Membership Control Limits: let us develop an algorithm for constructing the fuzzy number and membership control limits for the same number $m (=10)$ of Markov dependent samples drawn from a production line, size of the k^{th} sample $\mathbf{x}_k = (X_{k1}, X_{k2}, X_{k3}, X_{k4})$ is $\sum_{j=1}^4 X_{kj} = n_k$ using the same membership functions describing the four linguistic terms of the product quality given in (1).

Algorithm using fuzzy mathematics:

Step1: Calculate the fuzzy mean $M_k = \sum_{j=1}^4 \pi_{kj} r_j$ and fuzzy standard deviation

$$SD_k = \sqrt{\sum_{j=1}^4 \pi_{kj} (r_j - M_k)^2} \text{ of the } k^{\text{th}} \text{ sample for } k = 1, 2, \dots, m$$

Step2: Calculate the grand mean $M = \frac{1}{m} \sum_{k=1}^m M_k$ and the Mean of standard deviations (MSD)

$$MSD = \delta = \frac{1}{m} \sum_{k=1}^m SD_k .$$

Step3: Centre Limit/Line (CL) is M which is a fuzzy set with membership function

$$\mu_{CL}(x) = \begin{cases} \frac{x}{M} & \text{for } 0 \leq x \leq M \\ \frac{1-x}{1-M} & \text{for } M \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Step4: Fix a known distance 'K' from the CL for monitoring the variations of the product quality under study like that of the Shewhart control chart.

Step5: Compute the membership function based LCL and the UCL as
Membership LCL = Max{0, (CL - K δ)} and Membership UCL = Min{1, (CL + K δ)}.

Step6: Compute the coefficient of deviation K of the grand mean by using the Monte Carlo simulation assuming that the type I error is prefixed.

Step7: Plot the data points $\{M_k; k=1, 2, \dots, m\}$ and draw the lines LCL, CL, and UCL to check if any of the data points fall outside the control lines. (This result is called the fuzzy membership control chart for monitoring the product quality under fuzzy mode transformation).

The computational details about mean M (=CL=0.345) and the mean of standard deviation MSD is provided in Table 4 for the 10 Markov dependent samples under investigation and when K=0.29, UCL, LCL values have also been computed using the above algorithm to draw the fuzzy control chart of Figure 4.

Table 4: Data points M_k for $k=1,2,\dots,10$

k	std	Sc	Tc	Fc	N_k	M_k	SD_k
1	39	39	29	13	120	0.311	0.304
2	14	57	35	26	132	0.435	0.315
3	43	48	27	25	143	0.353	0.344
4	57	29	29	10	125	0.250	0.294
5	32	26	35	17	110	0.371	0.330
6	16	47	16	16	95	0.375	0.315
7	49	25	49	19	142	0.348	0.328
8	51	39	43	21	154	0.340	0.329
9	13	50	12	25	100	0.438	0.348
10	73	35	31	11	150	0.233	0.287
sum						3.453	3.193
M=							MSD=
						0.345	0.319

To compute the membership function based LCL and the UCL as in Step5 of the above algorithm using the formulae 'Membership LCL= $\text{Max}\{0, (CL- K\delta)\}$ ' and Membership UCL= $\text{Min}\{1, (CL+K \delta)\}$, K is selected such that both the LCL and the UCL of the FCC for all periods and the TPCC of the first period (as it belongs to the in-control period of the process under study) are identical and thus it is estimated that $K=0.2744$.

$$\text{Membership LCL} = \text{Max}\{0, (CL- K\delta)\} = \text{Max}\{0, (345- K0.319)\} = 0.2575$$

and

$$\text{Membership UCL} = \text{Min}\{1, (CL+ K\delta)\} = \text{Max}\{0, (345+K0.319)\} = 0.4325.$$

Inspecting the M_k values that are displayed in the Fuzzy control chat of Figure 4, it is found that corresponding to sample mean values M_k of sample numbers $k=2$, 4, 9 and $k=10$ fall outside the UCL/LCL in which the sample mean corresponding to

$k=9$ is not found as a point outside the UCL/LCL in the transition probability control chart of Figure 3. This fact leads to conclusion that FCC performs better than the TPCC provided the sample point of the $k=4^{th}$ period is a true alarm due to some assignable cause; otherwise the TPCC is performing as efficient as the FCC.

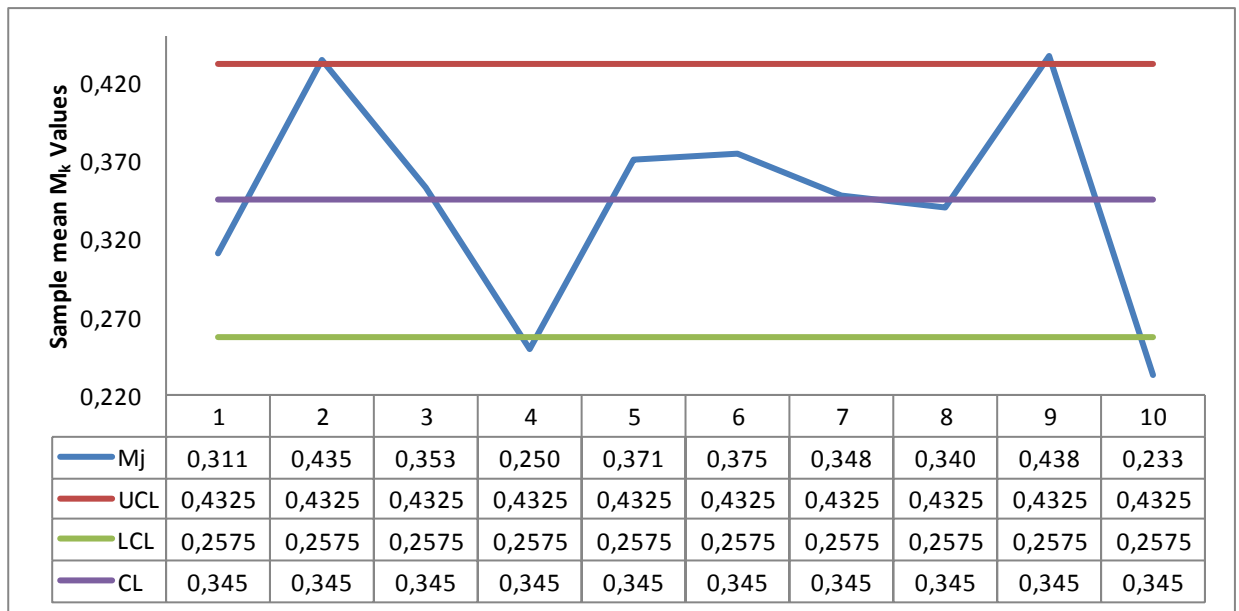


Figure 4: Fuzzy Control Chart

5 Concluding remarks

A solution to the problem of drawing adequately large samples sizes when the output rate of the production process is small for the construction of attribute control charts is provided through the formulation of the MC defined on the state space $\mathbb{S} = \{1=High\ Standard(LT_1), 2=Medium\ Standard(LT_2), 3=Low\ Standard(LT_3), 4=Poor\ Standard(LT_4)\}$ to include at least one item in each category of $\{LT_j\}$ such that the normality assumption is not violated. Further, this paper examines the uses of selecting Markov dependent small samples and the stationary distribution of that underlying MC on the state space of linguistic terms \mathbb{S} . Since the states associated

with such linguistic terms of the production process change randomly, it is generally impossible to predict the exact state of the production process in the future. However, through the stationary distribution of the TPM of the MC, it also explains how to generate expected larger sample sizes by assigning a suitable larger size at random to each of the observed small sample. For these larger sized samples, three different control charts viz., TPCC, generalized 'p' chart, and the FCC have been constructed. Data points corresponding to sample-2 and sample-9 fall either outside the UCL or on its border in each of these control charts. Additionally, sample number $k=10$ from TPCC and sample numbers $k=4$ and $k=10$ from FCC fall below the LCL. These facts lead to conclusion that FCC performs better than the TPCC provided the sample point of the $k=4^{\text{th}}$ period is a true alarm due to some assignable cause; otherwise the TPCC is performing as efficient as the FCC.

ACKNOWLEDGEMENT: Both authors thank the authorities of University of Botswana for their moral supports.

References

- [1]C.B. Cheng, Fuzzy process control: construction of control charts with fuzzy numbers, *Fuzzy Sets and Systems*, **154** (2), (2005), 287-303.
- [2]W.G. Cochran, Some methods for strengthening the common χ^2 tests, *Biometrics*, **10**(4), (1954), 417–451.
- [3]K. Demirli and S. Vijayakumar, Fuzzy logic based assignable cause diagnosis using control chart patterns, *Information Sciences*, **180**, (2010), 3258- 3272.
- [4]A. Faraz, R.B. Kazemzadeh, M.B. Moghadam and A. Bazdar, Constructing a fuzzy Shewhart control chart for variables when uncertainty and randomness are combined, *Journal of Quality & Quantity*, **44**(5), (2009).905-914.

- [5] M. Gülbay, C. Kahraman and D. Ruan, α - Cuts fuzzy control charts for linguistic data, *International Journal of Intelligent Systems*, **19**, (2004), 1173-1196.
- [6] R.F. Hamid and Pooyan Fekraty, Comparing Fuzzy Charts with Probability Charts and Using Them in a Textile Company, *The Journal of Mathematics and Computer Science*, **1**(4), (2010), 258-272.
- [7] O. Hryniewicz, Statistics with fuzzy data in statistical quality control, Soft Computing-A Fusion of Foundations, *Methodologies and Applications*, **12**(3), (2007), 229-234.
- [8] A. Kanagawa, F. Tamaki and H. Ohta, Control charts for process average and variability based on linguistic data, *Intelligent Journal of Production Research*, **31**(4), (1993), 913-922.
- [9] S. Karlin and Taylor, H.M, *A First Course in Stochastic Processers*, Academic Press, New York, 2nd ed., 1975.
- [10] J. Medhi, *Stochastic processes*. New Age international Pub., New Delhi, 2nd ed., 1994.
- [11] D.C. Montgomery, *Introduction to Statistical Quality Control*, 5th edition, John Wiley & Sons Inc., NY, USA., 1996.
- [12] T. Raz, and J.H. Wang, Probabilistic and memberships approaches in the construction of control charts for linguistic data, *Production Planning and Control*, **1**, (1990), 147-157.
- [13] M.H. Shu and H.C. Wu, Fuzzy X and R control charts: Fuzzy dominance approach, *Computers & Industrial Engineering*, **613**, (2011), 676-686.
- [14] R. Sivasamy and S. Jayanthi, Charts for detecting Small Shifts, *Economic Quality Control*, Hildermann Verlag, **18**(1), (2003), 91-100.
- [15] S. Sorooshian, Basic developments of Quality Characteristics Monitoring, *Journal of Applied Mathematics*, **2013**, (2013), 1-8.
- [16] H. Taleb and M. Limam, On fuzzy and probabilistic control charts, *International Journal of Production Research*, **40**(12), (2002), 2849 - 2863.

- [17]J.H. Wang. and T. Raz, on the construction of control charts for linguistic variables, *The International Journal of Production Research*, **28**(3), (1990), 477-487.
- [18]L.A. Zadeh, Fuzzy sets, *Information and Control*, **8**, (1965), 338-353.
- .