Journal of Mathematics and Statistics 4 (1): 26-31, 2008
ISSN 1549-3644

Construction of Control Charts Using Fuzzy Multinomial Quality

V. Amirzadeh, M. Mashinchi, M.A. Yaghoobi
Department of Statistics, Faculty of Mathematics and Computer,
Shahid Bahonar University of Kerman, Kerman, Iran.

Abstract: Control charts are the simplest type of on-line statistical process control techniques. One of the basic control charts is p-chart. In classical p-charts, each item classifies as either "nonconforming" or "conforming" to the specification with respect to the quality characteristic. In practice, one may classify each item in more than two categories such as "bad", "medium", "good", and "excellent". Based on this, we introduce a fuzzy multinomial chart (FM-chart) for monitoring a multinomial process. Control limits of FM-chart are obtained by using the multinomial distribution and the degrees of membership which one assigned to the distinct categories. The comparison of the FM-chart and the p-chart based on a food production process illustrates that the FM-chart leads to better results.

Keywords: Fuzzy multinomial control chart, p-chart, Linguistic variable, Membership function, Fuzzy statistics

INTRODUCTION

Control charts are widely used for monitoring and examining a production process. The power of control charts lies in their ability to detect process shifts and to identify abnormal conditions in the process. This makes possible the diagnosis of many production problems and often reduces losses and brings substantial improvements in product quality. In 1924, Walter Shewhart designed the first control chart and proposed a general model for control charts as follows: Let w be a sample statistic that measures some quality characteristic of interest. Moreover, suppose that the mean of w is μ_w and the standard deviation of w is σ_w. Then the center line (CL), the upper control limit (UCL), and the lower control limit (LCL) are defined as follows:

$$\begin{cases} UCL = \mu_w + k\sigma_w \\ CL = \mu_w \\ LCL = \mu_w - k\sigma_w \end{cases} \qquad \textbf{(1)}$$

where k is the "distance" of the control limits from the center line, expressed in standard deviation units.

Control charts are constructed and operated with data collected from the process. The data collected should represent the various levels of the quality characteristic associated with the product. The characteristics might be measurable on numerical scales, such as length, weight, voltage, etc., in which case control charts for variables are used. These included the $\overline{X}$-chart for controlling the process average and the R-chart (or S-chart) for controlling the process variability. If the quality-related characteristics cannot be represented in numerical form, such as characteristics for appearance, softness, colour, etc., then control charts for attributes are used [4]. Product units are classified either as "conforming" or "nonconforming", depending upon whether or not they meet specification. The p-chart is used to monitor the fraction nonconforming units. In p-chart, control limits calculate by using the normal approximation.

Linguistic scales are commonly used in industry to express properties or characteristic of products. Typically, the conformity to specifications on a quality standard is evaluated onto a two-state scales, e.g. acceptable or unacceptable, good or bad, and so on. However the binary classification might not be

Corresponding Author: Vahid Amirzadeh, Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman. Iran, Tel & Fax: +98 341 3220057,

appropriate in many situations, where product quality can assume more intermediate states. The assignment of weights, to reflect the degree of severity of product nonconformity has been adopted in many circumstances. Different numbers of weights are assigned to each class and the total number of weights is monitored with some control charts for defectives [9]. This approach requires the ability to classify each state uniquely into one of several mutually exclusive classes with well-defined boundaries between them. Quite often, there is some vagueness in the judgment applied by human quality inspectors, especially when dealing with characteristics that are evaluated subjectively. The vagueness present in linguistic variables may be treated formally with the help of fuzzy set theory.

Zadeh [10] In 1965, introduced the notion of fuzzy sets. After that, there have been efforts to apply it in statistics [7,8]. When products are classified into mutually exclusive linguistic categories, fuzzy control charts are used. Different procedures are proposed to construct these charts. Raz and Wang [5,9] developed fuzzy control charts for linguistic data which are mainly based on membership and probabilistic approaches. Kanagawa *et al.*[3] proposed an assessment of intermediate quality levels instead of the traditional binary judgment. Gulby *et al.* [2] proposed α-level fuzzy control charts for attributes in order to reflect the vagueness of data and tightness of inspection. This work attempts to construct a new fuzzy multinomial control chart (namely FM-chart) for linguistic variables. To this end, the control limits of the FM-chart are introduced. The FM-chart is able to deal with a linguistic variable which is classified in more than two categories. Therefore the FM-chart provides more information than p-chart. This fact is illustrated by an example from a production process.

MATERIALS AND METHODS

A linguistic variable differs from a numerical variable in that its values are not numbers but words or phrases in some language. In the context of fuzzy set theory, a linguistic variable $\tilde{L}$ is characterized by its term set $\{l_1, l_2, ..., l_k\}$, which is the set of all possible values that $\tilde{L}$ may take on. Each term l_i has a weight $(\tilde{L}(l_i))$ to reflect its degree of membership in the set. This can be denoted by a fuzzy set as $\tilde{L} = \{(l_1, \tilde{L}(l_1)), (l_2, \tilde{L}(l_2)), ..., (l_k, \tilde{L}(l_k))\}$, [11].

For example, on a production line, a visual control of the corking and closing process might have the following assessment possibilities [1]:

1. "reject" if the cork does not work;
2. "poor quality" if the cork works but has some defects;
3. "medium quality" if the cork works and has no defects, but it has some aesthetic flaws;
4. "good quality" if the cork works and has no defects, but has only a few aesthetic flaws;
5. "excellent quality" if the cork works and has neither defects nor aesthetic flaws of any kind.

The monitoring of production, using a sampling control technique, is aimed at recognizing and, possibly, correcting unfavorable trends and out of control conditions. In order to do this, the five classifications listed above could have different degrees of membership. For example, one may assign to the five quality levels 1-5, the degrees of membership: 1, 0.75, 0.5, 0.25, and 0, respectively. In other words, if linguistic variable $\tilde{L}$ be "the quality of the corking and closing process", then $\tilde{L}$ ={(reject,1), (poor quality, 0.75), (medium quality, 0.5), (good quality, 0.25), (excellent quality, 0)}. Although the numerical conversion of verbal information simplifies subsequent analysis, it also gives rise to two problems.

The first is concerned with the validity of encoding a discrete verbal scale into a numerical form. This approach introduces properties that were not present in the original linguistic scale (for example, is it legitimate to assume that the difference between the "reject" state and the "poor quality" state is the same as that between "medium" and "good quality" states?). The second is related to the absence of consistent criteria to select the values of the degree of membership. It is obvious that changing the values of the degree of membership may determine a change in obtained results [1, 6]. In order to minimize these problems, it is recommended that the number of categories with their degrees of membership be arrived at after discussion with experts on the process concerned.

RESULTS AND DISCUSSION

Fuzzy multinomial control chart: In the following, we propose a new approach for construction of a control chart. The statistical principles underlying the fuzzy multinomial control chart (FM-chart) are based on the multinomial distribution.

Let $\tilde{L} = \{(l_1, \tilde{L}(l_1)), (l_2, \tilde{L}(l_2)), ..., (l_k, \tilde{L}(l_k))\}$ be a linguistic variable. In addition suppose that the production process is operating in a stable manner, and p_i is the probability that an item is l_i, $i = 1,2,...k.$ Moreover, successive items produced are independent. Assume that a random sample of n items of the product is selected. Let X_i, $i = 1,2,...,k$, be the number of items of the product that are l_i, $i = 1,2,...k.$ Then $(X_1, X_2, ..., X_k)$ has a multinomial distribution with parameters n and $p_1, p_2, ..., p_k$. It is known that each X_i, $i = 1,2,...,k$, marginally has a binomial distribution with the mean np_i and variance $np_i(1 - p_i)$, $i = 1,2,...,k$, respectively. Now assign the degree of membership $\tilde{L}(l_i)$ to each item in the ith category, and define the weighted average of $\tilde{L}(l_i)$, $i = 1,2,...,k$, denoted by $\bar{\tilde{L}}$ as follows:

$$\bar{\tilde{L}} = \frac{\sum_{i=1}^{k} X_i \tilde{L}(l_i)}{\sum_{i=1}^{k} X_i} = \frac{\sum_{i=1}^{k} X_i \tilde{L}(l_i)}{n}.$$

We introduce the control limits for FM-chart as:

$$\begin{cases} UCL = E(\bar{\tilde{L}}) + k\sqrt{Var(\bar{\tilde{L}})}, \\ CL = E(\bar{\tilde{L}}), \\ LCL = E(\bar{\tilde{L}}) - k\sqrt{Var(\bar{\tilde{L}})}, \end{cases} \quad (2)$$

where k (usually k =3) is the "distance" of the control limits from the center line. The following theorem shows how to compute $E(\bar{\tilde{L}})$ and $Var(\bar{\tilde{L}})$.

Theorem 1:
Let $\tilde{L} = \{(l_1, \tilde{L}(l_1)), (l_2, \tilde{L}(l_2)), ..., (l_k, \tilde{L}(l_k))\}$ be a linguistic variable such that p_i is the probability that an item is l_i, $i = 1,2,...,k$. Suppose that a random sample of n items of the product is selected. Let X_i, $i = 1,2,...,k$, be the number of items of the product that are l_i, $i = 1,2,...,k$, then:

$$i)\ E(\bar{\tilde{L}}) = \sum_{i=1}^{k} p_i \tilde{L}(l_i)$$

$$ii)\ Var(\bar{\tilde{L}}) = \frac{\sum_{i=1}^{k} p_i(i - p_i)\tilde{L}^2(l_i) - 2\sum_{i=1}^{k}\sum_{i<j}\sum_{j=1}^{k} p_i p_j \tilde{L}(l_i)\tilde{L}(l_j)}{n} \quad (3)$$

Proof: X_i, $i = 1,2,...,k$ has a binomial distribution with the mean np_i and variance $np_i(1 - p_i)$, $i = 1,2,...,k$, respectively,

and $Cov(X_i, X_j) = -p_i p_j$, $i \neq j.$ Therefore:

$$i)\ E(\bar{\tilde{L}}) = \frac{\sum_{i=1}^{k} \tilde{L}(l_i) E(X_i)}{n} = \frac{\sum_{i=1}^{k} \tilde{L}(l_i) np_i}{n} = \sum_{i=1}^{k} p_i \tilde{L}(l_i)$$

$$ii)\ Var(\bar{\tilde{L}}) = \frac{\sum_{i=1}^{k} \tilde{L}^2(l_i) Var(X_i) - 2\sum_{i=1}^{k}\sum_{i<j}\sum_{j=1}^{k} \tilde{L}(l_i)\tilde{L}(l_j) Cov(X_i, X_j)}{n}$$

$$= \frac{\sum_{i=1}^{k} \tilde{L}^2(l_i) np_i(1 - np_i) - 2\sum_{i=1}^{k}\sum_{i<j}\sum_{j=1}^{k} \tilde{L}(l_i)\tilde{L}(l_j) np_i p_j}{n}$$

$$= \frac{\sum_{i=1}^{k} p_i(i - p_i)\tilde{L}^2(l_i) - 2\sum_{i=1}^{k}\sum_{i<j}\sum_{j=1}^{k} p_i p_j \tilde{L}(l_i)\tilde{L}(l_j)}{n}$$

Remark 1: If $\tilde{L} = \{(l_1, 1), (l_2, 0)\}$ is a linguistic variable, then the FM-chart reduces to a p-chart with $p = P_r$(an item is l_1).

An illustrative example: In food process industry, packaging of a frozen food is important quality characteristic that has to be monitored [4]. The product item's packaging, may be classified by an expert team as either "excellent", "good", "medium" or "bad" with the degrees of membership 0, 0.25, 0.5, and 1, respectively. For control of the quality packging process, 30 samples of size 50 are selected. The data with $\tilde{\bar{L}}_i$ and $\hat{p}_i$ are given in Table 1.

Table 1. The data of samples of size 50

i	Bad	Medium	Good	Excellent	$\tilde{\bar{L}}_i$	$\hat{p}_i$
1	8	9	25	8	0.375	0.16
2	10	8	25	7	0.405	0.20
3	6	8	28	8	0.340	0.12
4	6	8	28	8	0.340	0.12
5	3	9	28	10	0.290	0.06
6	5	9	27	9	0.325	0.10
7	10	8	24	8	0.400	0.20
8	6	9	27	8	0.345	0.12
9	9	9	26	6	0.400	0.18
10	7	8	27	8	0.355	0.14
11	3	9	30	8	0.300	0.06
12	4	10	29	7	0.325	0.08
13	11	9	24	6	0.430	0.22
14	7	10	26	7	0.370	0.14
15	16	9	20	5	0.510	0.32
16	5	9	28	8	0.330	0.10
17	7	9	27	7	0.365	0.14
18	3	9	30	8	0.300	0.06
19	9	9	26	6	0.400	0.18
20	7	8	27	8	0.355	0.14
21	12	15	20	3	0.490	0.24
22	11	11	22	6	0.440	0.22
23	17	9	20	4	0.530	0.34
24	10	9	24	7	0.410	0.20
25	6	9	28	7	0.350	0.12
26	7	10	25	8	0.365	0.14
27	5	12	25	8	0.345	0.10
28	9	8	26	7	0.390	0.18
29	5	12	25	8	0.345	0.10
30	4	9	28	9	0.310	0.08

Suppose that the process is in control in the period corresponding to first ten samples. The sample proportions for the base period estimate as follows:

$$p_B = 0.14, \quad p_M = 0.17, \quad p_G = 0.53, \quad p_E = 0.16.$$

Hence, the central line and control limits are:

$$CL = 0.3575, \quad UCL = 0.4832, \quad LCL = 0.2318.$$

Now, the $\tilde{\bar{L}}_i$ for the 30 samples are plotted in an FM -chart (Figure 1).

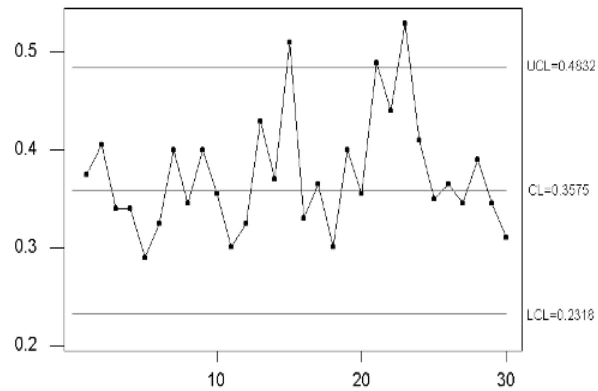

Fig. 1: FM -chart of the $\tilde{\bar{L}}_i$ for the 30 samples

Figure 1 shows that the samples 15, 21, and 23 are out of control. In the p-charts, the center line represents the fraction of nonconforming items. For this example, we assume that only all the items classified as "Bad" are nonconforming. Therefore the central line and control limits for a p-chart are:

$$CL = 0.14, \quad UCL = 0.2872, \quad LCL = 0.$$

The p -chart for the 30 samples is shown in Figure 2.

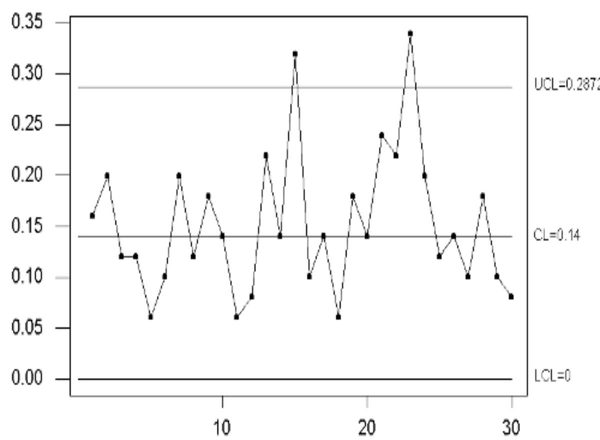

Fig. 2: The p -chart for the 30 samples

As it can be seen from Figure 2, only the samples 15 and 23 are out of control. A p -chart compares with the corresponding FM -chart on the basis of the samples out of control and, the analysis of the probability of type II errors.

In the FM -chart, three samples (15, 21, and 23) are out of control whereas in the p -chart, only samples 15 and 23 are out of control. Since an FM -chart utilizes more information, we expect this chart to perform better than a p -chart for linguistic data. In order to compute the type II errors at the values in samples 15 and 23, we suppose that the process shifts from the in control values (i.e., $H_0: \ p_B = 0.14, p_M = 0.17, p_G = 0.53, p_E = 0.16)$ to these sample values (i.e., $H_1: p_B = 0.32, \ p_M = 0.18, \ p_G = 0.4, \ p_E = 0.1$ For sample 15, and $H_1': \ p_B' = 0.34, \ p_M' = 0.18, p_G' = 0.4, \ p_E' = 0.08$ for sample 23).The corresponding computed probability of type II errors in the p -chart and the FM -chart are shown in Table 2.

Table 2: The probability of type II errors

	H_1	H_1'
p -chart	0.33	0.23
FM -chart	0.3	0.18

It can be seen that the probability of type II errors, under H_1 and H_1' for the FM -chart are smaller than the p -chart. Moreover, since p -chart depends only on the value of p_B, it will not register change if p_B remain constant even though p_M, p_G, and p_E vary. However, since $\overline{\overline{L}}$ is a function of all four probabilities ($p_B, p_M, \ p_G$, and p_E), the FM -chart will register a variation in any one of these. Consequently, we conclude that the FM -chart leads to better results than the p -chart, if the number of categories and their degrees of membership are selected well.

CONCLUSION

When quality control by variables is not feasible, linguistic data provides more information than the binary classification used in control by attributes. The representation of linguistic variable as fuzzy set, retains the ambiguity and vagueness inherent in natural languages and improves the expressive ability of quality assurance inspectors. In this paper, we have attempted to extend the use of control charts to linguistic variables by presenting an approach for determining the center line and control limits. The approach represents the linguistic variable as a fuzzy set. Then a fuzzy multinomial control chart (FM -chart) is introduced for monitoring a production process. An illustrative example from a production process is discussed to show the efficiency of the FM-chart. Some problems, however, still remain. First, how many linguistic terms should be defined? Second, how should the degrees of membership of linguistic terms be constructed? These problems will be the subjects of future research.

ACKNOWLEDGEMENTS

The first author would like to thank partially support of Fuzzy Systems and its Applications Center of Excellence, Shahid Bahonar University of Kerman, Iran.

REFERENCES

1. Franceschini, F. and D. Romano, 1999. Control chart for linguistic variables: A method based on the use of linguistic quantifiers. International Journal of Production Research. 37: 3791-3801.
2. Gulbay, M., C. Kahraman and D. Ruan, 2004. α-cut fuzzy control charts for linguistic data. International Journal of Intellgent Systems. 19: 1173-1196.
3. Kanagawa, A., F. Tamaki and H. Ohta, 1993. Control charts for process average and variability based on linguistic data. International Journal of Production Research. 31: 913-922.
4. Montgomery, D.C., 2005. Introduction to Statistical Quality Control (5th ed.). John Wiley and Sons. New York.
5. Raz, T. and J.H. Wang, 1990. Probabilistic and membership approaches in the construction of control charts for linguistic data. Production Planning and Control. 1: 147-157.

6. Steiner, S.F., P.L. Geyer and G.O. Wesolowsky, 1994. Control charts based on grouped observations. International Journal of Production Research. 32: 75-91.
7. Taheri, S.M., 2003. Trends in fuzzy statistics Austrian Journal of Statistics. 32: 239-257.
8. Viertl, R., 1996. Statistical Methods for Non-precise data. CRC Press. Boca Raton. Florida.
9. Wang, J.H. and T. Raz, 1990. On the costruction of control charts using linguistic data. International Journal of Production Research. 28: 477-487.
10. Zadeh, L.A., 1965. Fuzzy sets. Information and control. 8: 338-359.
11. Zimmermann, H.J., 1996. Fuzzy Set Theory and Its Applications. Kluwer Academic Publishers. Boston.