A CUSUM Control Chart for Fuzzy Quality Data

Dabuxilatu Wang

School of Mathematics and Information Sciences,
Guangzhou University, No. 230 Wai-Huan West Road ,
University Town, Panyu District, Guangzhou, 510006, P.R.China
dbxlt0@yahoo.com

Summary. Based on the concept of fuzzy random variables, we propose an optimal representative value for fuzzy quality data by means of a combination of a random variable with a measure of fuzziness. Applying the classical Cumulative Sum (CUSUM) chart for these representative values, an univariate CUSUM control chart concerning *LR*-fuzzy data under independent observations is constructed.

Key words: *statistical process control; Cumulative sum chart; representative values; fuzzy sets.*

1 Introduction

Cumulative Sum (CUSUM) control chart proposed by Page [9] is a widely used tool for monitoring and examining modern production processes. The power of CUSUM control chart lies in its ability to detect small shifts in processes as soon as it occurs and to identify abnormal conditions in a production process. For example, for a given sequence of observations $\{X_n, n = 1,2,...\}$ on normal population, the monitored parameter of interest is typically the process mean, $\mu_n = E(X_n)$, the purpose is to detect a small change in the process mean, one might specifies the levels μ_0 and $\mu_1 > \mu_0$ (or $\mu_1 < \mu_0$) such that under normal conditions the values of μ_n should fall below (or above) μ_0 and the values of μ_n above (or below) μ_1 are considered undesirable and should be detected as soon as possible. The CUSUM chart can be used to monitor above process with the test-statistics $S_n = \max\{0, S_{n-1} + X_n - K\}$ (or $T_n = \min\{0, T_{n-1} + X_n + K\}$) and signal if $S_n > b$ (or $T_n < -b$), where b is the control limit derived from a confidence interval assuming a Gaussian distributed observation, X_n $(n \geq 1)$ are the sample means at time t_n and $S_0 = T_0 = 0$, and K is the reference value. It is well-known that CUSUM chart is more sensitive than Shewhart chart ($\overline{X}$-chart) for detecting certain small changes in process parameters.

The random processes encountered in industrial, economic, etc. are typically quality monitoring processes. J.M. Juran, an authority in international quality control circles, has pointed out that quality to customers, is its suitability rather than its

D. Wang: *A CUSUM Control Chart for Fuzzy Quality Data*, Advances in Soft Computing **6**, 357–364 (2006)
www.springerlink.com

conformity to certain standards. End-users seldom know what standards are. Customers' appraisal on quality is ways based on whether the products they have bought are suitable or not and whether that kind of suitability will last [1].

This is a kind of quality outlook which attaches primary importance to customers' feeling, so vague attribute of quality appraisal criterion and appraising customers' psychological reactions should, by no means, be ignored.

The fuzzy set theory [16] may be an appropriate mathematical tool for dealing with vagueness and ambiguity of the quality attribute. So it is very natural to introduce the concept of fuzzy set to the concept of quality and thus fuzzy quality is formed. As regards fuzzy quality, its "suitability" quality standard is expressed in the form of a fuzzy set. Also an outcome of the observation on quality characteristics may be appropriately represented by a fuzzy set because it is difficult to obtain a precise quality description of the inspected item in some case.

There are some literature on constructions of control charts based on fuzzy observations by [14], [11], [6], [4, 5], [13], [12] and [2]. Basically, the works mentioned above include two kinds of controlling methods , one of which is utilizing probability hypotheses testing rule for the representative values of fuzzy quality data, and the other is using a soft control rule based on possibility theory. However, how to deal with optimally both randomness and fuzziness of the process quality data is still a problem.

2 *LR*-fuzzy Data and the Measure of Fuzziness

2.1 *LR*-fuzzy Data

A fuzzy set on $\mathbb{R}$, the set of all real numbers, is defined to be a mapping $u : \mathbb{R} \to [0,1]$ satisfying the following conditions:

(1) $u_\alpha = \{x|u(x) \geq \alpha\}$ is a closed bounded interval for each $\alpha \in (0,1]$.
(2) $u_0 = supp$u is a closed bounded interval.
(3) $u_1 = \{x|u(x) = 1\}$ is nonempty.

where supp$u = cl\{x|u(x) > 0\}$, cl denotes the *closure* of a set. Such a fuzzy set is also called a *fuzzy number*. The following parametric class of fuzzy numbers, the so-called *LR*-fuzzy numbers, are often used in applications:

$$u(x) = \begin{cases} L(\frac{m-x}{l}), & x \leq m \\ R(\frac{x-m}{r}), & x > m \end{cases}$$

Here $L : \mathbb{R}^+ \to [0,1]$ and $R : \mathbb{R}^+ \to [0,1]$ are fixed left-continuous and non-increasing function with $L(0) = R(0) = 1$. L and R are called left and right shape functions, m the central point of u and $l > 0$, $r > 0$ are the left and right spread of u. An *LR*-fuzzy number is abbreviated by $u = (m,l,r)_{LR}$, especially $(m,0,0)_{LR} := m$. Some properties of *LR*-fuzzy numbers for operations are as follows:

$$(m_1,l_1,r_1)_{LR}+(m_2,l_2,r_2)_{LR} = (m_1+m_2,l_1+l_2,r_1+r_2)_{LR}$$
$$a(m,l,r)_{LR} = \begin{cases} (am,al,ar)_{LR}, & a>0 \\ (am,-ar,-al)_{RL}, & a<0 \\ 0, & a=0 \end{cases}$$
$$(m_1,l_1,r_1)_{LR}-m_2 = (m_1-m_2,l_1,r_1)_{LR}$$

For further properties of *LR*-fuzzy numbers the readers are refereed to [3].

Let $L^{(-1)}(\alpha) := \sup\{x \in R | L(x) \geq \alpha\}, R^{(-1)}(\alpha) := \sup\{x \in R | R(x) \geq \alpha\}$. Then for $u = (m,l,r)_{LR}$, $u_\alpha = [m - lL^{(-1)}(\alpha), m + rR^{(-1)}(\alpha)]$, $\alpha \in [0,1]$.

An useful approach has been summarized by Cheng [2] for generating a fuzzy number based on a group experts' scores on a fuzzy quality item in a quality control process. By this approach, we may assign a fuzzy number for each outcome of a fuzzy observation on quality monitoring process. In this paper, we assume that the quality data collected from the fuzzy observation process can be assigned *LR*-fuzzy numbers. Such data is also called *LR*-fuzzy data. For example, the color uniformity of a TV set [1] under user's suitability quality view is with a fuzzy quality standard which could be expressed in a form of *LR*-fuzzy data $(d_0,5,5)_{LR}$, where $L(x) = R(x) = \max\{0, 1-x\}$ is the shape function of a triangular fuzzy number, and d_0 is the designed value of the color uniformity. For the operational simplicity and a better description of fuzziness for fuzzy quality items, the triangular fuzzy number are often used.

LR-fuzzy random variable $X = (m,l,r)_{LR}$ has been defined by Körner [7], where m,l,r are three independent real-valued random variables with $P\{l \geq 0\} = P\{r \geq 0\} = 1$. Considering the fuzzy observations on a quality monitoring process, it is obvious that the *LR*-fuzzy data can be viewed as realizations of an *LR*-fuzzy random variable. Assuming the observational distribution is approximately normal, then the central variable m of an *LR*-fuzzy sample $X = (m,l,r)_{LR}$ obtained by method in [2] from the fuzzy observation process can be viewed as a Gaussian variable, and the spread variables l, r may be evenly distributed. The i^{th} sample X_i is assumed to be a group mean of size n_i, $\{(x_{i1},b_{i1},c_{i1})_{LR},\ldots,(x_{in_i},b_{in_i},c_{in_i})_{LR}\}$, i.e.

$$X_i = (\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, \frac{1}{n_i}\sum_{j=1}^{n_i} b_{ij}, \frac{1}{n_i}\sum_{i=1}^{n_i} c_{ij})_{LR} = (\bar{x}_i,\bar{b}_i,\bar{c}_i)_{LR},$$

and simply denoted by $X_i = (m_i,l_i,r_i)_{LR}$. By the central limit theorem, if the group size is relatively large, then l_i, r_i are approximately Gaussian variables.

2.2 The Measure of Fuzziness

The *LR*-fuzzy quality data are not easy to be plotted on a control chart directly. Therefore, it is necessary to convert a fuzzy data (a fuzzy sample) into a scalar (a real random variable) for plotting, such scalar (random variable) would be an optimal representative of the fuzzy data (the fuzzy sample). Some approaches for determining the representing value of a fuzzy data have been proposed in [14] and [6], etc.. In general, there are no absolute criteria for choosing methods on determining the representative value. However, we usually expect a method for determining one which is not only with lower complexity in computation but also with an optimal representativeness.

Recalling the concept of fuzzy random variables [7],[8], [10], we are aware of that fuzzy random variables are devoted to deal with the inherent randomness and fuzziness of samples simultaneously. Thus, we emphasize that a representative value should properly represent the main characteristics, randomness and fuzziness, of a fuzzy quality sample. Such features can be abstracted easier in the case of *LR*-fuzzy sample than that of other fuzzy sample because we are able to represent the randomness by the central variable and to represent the fuzziness of the fuzzy quality data by employing the concept of a measure of fuzziness.

A number used for measuring the fuzziness of a fuzzy set is a very important index when we deal with fuzzy concepts and fuzzy information. Fuzziness level of a fuzzy set is usually determined by the fuzziness level of each possible elements of the fuzzy set. For example, if the membership degree of one element is near 1, then the affirmation level with respect to the element must be high, and thus fuzziness level of the element becomes low; if the membership degree of one element is around 0.5, then its belongingness is extremely unsteady, and thus the fuzziness level of the element becomes high, and so on. Various measuring methods have been proposed based on the concept of measure of fuzziness [3] [17], for instance, Minkowski's measure of fuzziness $D_p(A)$ for a fuzzy set A on a discrete finite domain is as follows:

$$D_p(A) = \frac{2}{n^{1/p}}(\sum_{i=1}^{n} |A(x_i) - A_{0.5}(x_i)|^p)^{1/p},$$

where $p > 0$, $A_{0.5}$ denotes the 0.5-level set of the fuzzy set A , and $A_{0.5}(x)$ denotes the indicator of the non-fuzzy set $A_{0.5}$, i.e.,

$$A_{0.5}(x) := I_{A_{0.5}}(x) = \begin{cases} 1, & x \in A_{0.5} \\ 0, & x \notin A_{0.5} \end{cases}$$

When $p = 1$, $D_1(A)$ is said to be Hamming's fuzziness measure, and when $p = 2$, $D_2(A)$ is called Euclid's fuzziness measure. We employ an extension of Hamming's fuzziness measure to define a measure of fuzziness $D(X)$ for the *LR*-fuzzy quality sample $X = (m,l,r)_{LR}$, i.e.

$$D(X) = \int_{-\infty}^{+\infty} |X(x) - X_{0.5}(x)| dx$$

Theorem 1. *Let $X = (m,l,r)_{LR}$ be a fuzzy quality sample, then it holds that*

$$\begin{aligned} D(X) &= l\left[L^{(-1)}(0.5) + \int_{L^{(-1)}(0.5)}^{L^{(-1)}(0)} L(x)dx - \int_0^{L^{(-1)}(0.5)} L(x)dx\right] \\ &+ r\left[R^{(-1)}(0.5) + \int_{R^{(-1)}(0.5)}^{R^{(-1)}(0)} R(x)dx - \int_0^{R^{(-1)}(0.5)} R(x)dx\right]. \end{aligned}$$

Proof. It is obvious that

$$X_{0.5}(x) = \begin{cases} 1, & x \in \left[m - lL^{(-1)}(0.5), m + rR^{(-1)}(0.5)\right] \\ 0, & x \notin \left[m - lL^{(-1)}(0.5), m + rR^{(-1)}(0.5)\right] \end{cases}$$

$$\begin{aligned} D(X) &= \int_{-\infty}^{+\infty} |X(x) - X_{0.5}(x)| dx \\ &= \int_{m-lL^{(-1)}(0.5)}^{m} (1 - L(\frac{m-x}{l}))dx + \int_{m}^{m+rR^{(-1)}(0.5)} (1 - R(\frac{x-m}{r}))dx \\ &+ \int_{m-lL^{(-1)}(0)}^{m-lL^{(-1)}(0.5)} L(\frac{m-x}{l})dx + \int_{m+rR^{(-1)}(0.5)}^{m+rR^{(-1)}(0)} R(\frac{x-m}{r})dx \\ &= l\left[L^{(-1)}(0.5) + \int_{L^{(-1)}(0.5)}^{L^{(-1)}(0)} L(x)dx - \int_{0}^{L^{(-1)}(0.5)} L(x)dx\right] \\ &+ r\left[R^{(-1)}(0.5) + \int_{R^{(-1)}(0.5)}^{R^{(-1)}(0)} R(x)dx - \int_{0}^{R^{(-1)}(0.5)} R(x)dx\right]. \end{aligned}$$

This completes the proof.

Example Let $u = (m_0, l_0, r_0)_{LR}$ be a triangular fuzzy data, where $L(x) = R(x) = \max\{0, 1 - x\}$, then

$$D(u) = \frac{l_0 + r_0}{4}.$$

3 Construction of a CUSUM Chart for *LR*-fuzzy Data

3.1 A Representative Value

We now define a representative value denoted by $Rep(X)$ for fuzzy sample $X = (m, l, r)_{LR}$ as follows:

$$Rep(X) = m + D(X)$$

Let

$$\beta_1 := L^{(-1)}(0.5) + \int_{L^{(-1)}(0.5)}^{L^{(-1)}(0)} L(x)dx - \int_{0}^{L^{(-1)}(0.5)} L(x)dx$$

$$\beta_2 := R^{(-1)}(0.5) + \int_{R^{(-1)}(0.5)}^{R^{(-1)}(0)} R(x)dx - \int_{0}^{R^{(-1)}(0.5)} R(x)dx$$

then

$$Rep(X) = m + l\beta_1 + r\beta_2$$

Here, the central variable m just represents the randomness of the *LR*- fuzzy quality sample $X = (m, l, r)_{LR}$ extremely, since by its membership $X(m) = 1$ it implies no fuzziness , and it also largely determine the location of the *LR*-fuzzy quality sample. On the other hand, random variable $l\beta_1 + r\beta_2$ properly represents the fuzziness level of the *LR*-fuzzy quality sample because it is derived from a standard measure of fuzziness of a fuzzy set, which is well defined with a theoretical supporting. A kind of combination of the randomness with fuzziness of the fuzzy quality sample is realized simply by arithmetic addition, thus the related computation for obtaining the representative value becomes much easier. For a given fuzzy quality data $u = (m_0, l_0, r_0)_{LR}$, then its representative value is $Rep(u) = m_0 + l_0\beta_1 + r_0\beta_2$, which is a fixed scalar. The present methods for calculating representative values in the case of *LR*-fuzzy quality data somewhat have an advantage over that proposed in [14] and [6]. For instance, calculating representative values were done in five ways in [14] and [6]: by using the fuzzy mode as $f_{mode} = \{x|A(x) = 1\}, \quad x \in [a,b]$; the α-level fuzzy midrange as $f_{mr}(\alpha) = \frac{1}{2}(\inf A_\alpha + \sup A_\alpha)$; the fuzzy median as f_{med}, which satisfies $\int_a^{f_{med}} A(x)dx = \int_{f_{med}}^b A(x)dx = \frac{1}{2}\int_a^b A(x)dx$; the fuzzy average as $f_{avg} = \int_a^b xA(x)dx / \int_a^b A(x)dx$;and the barycentre concerned with Zadeh's probability measure of fuzzy events as $Rep(A) = \int_{-\infty}^{\infty} xA(x)f(x)dx / \int_{-\infty}^{\infty} A(x)f(x)dx$. Where A is a fuzzy set on some interval $[a,b] \subset \mathbb{R}, a < b$. In general, the first two methods are easier to calculate than the last three as well as our method, however, they only took account of the randomness of the fuzzy sample , e.g. $f_{mode} = m$ when the fuzzy sample is $X = (m,l,r)_{LR}$, which obviously may lead to a biased result. The third method used a non-standard measure of fuzziness , thus the fuzzy median may also be a biased representative of a fuzzy sample. The last two methods are reasonable, but the representative values derived from the methods are not easy to calculate. We can easily check that our method is easier to calculate than the fuzzy average and barycentre methods in the case of *LR*-fuzzy quality sample. We would like to point out that the accuracy of the representative for the given fuzzy sample is more important for constructing a representing control chart devoted for monitoring fuzzy quality, an inaccurate representative will lead to more false alarm or a wrong control scheme deviated from the original reality of fuzzy data . Also it is a common sense that every fuzzy data is characterized by the both randomness and fuzziness. Our proposed representative value is considerably accurate and simply and very comprehensive because we fully take the randomness as well as fuzziness measured by a standard fuzziness measure into account.

3.2 Construction of a CUSUM Chart

Using the classical CUSUM interval scheme [15], we can design a corresponding chart for the representative values of *LR*-fuzzy quality samples. As that mentioned in Subsection 2.1, the spread variable of a sample might be evenly distributed though the central variable is Gaussian, we need sampling in groups of varying number n_i of observations each, and n_i is relatively large, for instance, $n_i \geq 25$. Let the observation is:

$$X_{ij} = (m_{ij}, l_{ij}, r_{ij})_{LR}, \quad i = 1,2,\ldots,k; j = 1,2,\ldots,n_i$$

then $x_{ij} := Rep(X_{ij}) = m_{ij} + l_{ij}\beta_1 + r_{ij}\beta_2$. The representative value of the samples mean (group mean), denoted by $\overline{x_i}$, can be worked out by the following two ways: (1). $\overline{x_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} = \overline{m_i} + \overline{l_i}\beta_1 + \overline{r_i}\beta_2$. (2).$\overline{X_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} = (\overline{m_i}, \overline{l_i}, \overline{r_i})_{LR}$, then $\overline{x_i} = Rep(\overline{X_i}) = \overline{m_i} + \overline{l_i}\beta_1 + \overline{r_i}\beta_2$. The standard error of mean for the representative values in group i, denoted by s_i, is:

$$s_i = \left(\frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left[(m_{ij} - \overline{m_i}) + (l_{ij} - \overline{l_i})\beta_1 + ((r_{ij} - \overline{r_i})\beta_2) \right]^2 \right)^{1/2}.$$

Then the standard error of samples mean can be estimated by

$$\hat{\sigma}_e = \left(\sum_{i=1}^{k} (n_i - 1)s_i^2 / \sum_{i=1}^{k} (n_i - 1) \right)^{1/2}.$$

Thus, we are able to construct a CUSUM control chart for the representative values of the samples as follows:

(1) Choose a suitable reference value T, here we assume that it is the overall mean $\hat{\mu}$ of the past observations.
(2) Use the standard scheme $h = 5, \quad f = 0.5$.
(3) Calculate the CUSUM S_n (Here S_n is with respect to the representative values of samples) with reference value $K_1 = T + f\hat{\sigma}_e = \hat{\mu} + 0.5\hat{\sigma}_e$. Keep it non-negative. Calculate the CUSUM T_n (Here T_n is with respect to the representative values of samples) with reference value $K_2 = T - f\hat{\sigma}_e = \hat{\mu} - 0.5\hat{\sigma}_e$. Keep it non-positive.
(4) Action is signalled if some $S_n \geq h\hat{\sigma}_e = 5\hat{\sigma}_e$ or some $T_n \leq -h\hat{\sigma}_e = -5\hat{\sigma}_e$.

This obtained CUSUM control chart is an appropriate representative CUSUM chart for the *LR*-fuzzy quality data involved process.

Conclusions

We have proposed an optimal representative value for fuzzy quality sample by means of a combination of a random variable with a measure of fuzziness. For *LR*- fuzzy data this kind of representative values are more accurate and easier to calculate, so by which the fuzzy control charts derived from using representative values methods could be improved to some sense. An accurate representative CUSUM chart for *LR*-fuzzy samples is preliminarily constructed. Likewise one could construct other control charts such as EWMA,P-chart and so on. The proposed representative value is expected to be extended to a general case where the normal, convex and bounded fuzzy quality data are monitored.

References

[1] Cen Y, 1996, Fuzzy quality and analysis on fuzzy probability, Fuzzy sets and systems, 83, 283-290.

[2] Cheng CB, 2005, Fuzzy process control: construction of control charts with fuzzy numbers, Fuzzy sets and systems, 154, 287-303.
[3] Dubois D, Prade H, 1980, Fuzzy sets and systems: Theory and Applcation. New York: Academic press.
[4] Grzegorzewski P, 1997, Control charts for fuzzy data, Proc. Fifth European congress on intelligent techniques and soft computing EUFIT'97, Aachen, pp. 1326-1330.
[5] Grzegorzewski P, Hryiewicz O, 2000, Soft methods in statistical quality control, Control and cybernet. 29, 119-140.
[6] Kanagawa A, Tamaki F, Ohta H, 1993, Control charts for process average and variability based on linguistic data, International Journal of Production Research, 2, 913-922.
[7] Körner R, 2000, An asymptotic α-test for the expectation of random fuzzy variables, Journal of Statistical Planning and Inferences. 83, 331-346.
[8] Kruse R, Meyer KD, 1987, Statistics with vague data. D. Reidel Publishing Comany.
[9] Page ES, 1954, Continuous inspection schemes, Biometrika, 41, 100-114.
[10] Puri MD, Ralescu D, 1986, Fuzzy Random Variables, Journal of Mathematical Analysis and Applications, 114, 409-422.
[11] Raz T, Wang JH, 1990, Probabilistic and membership approaches in the construction of control charts for linguistic data, Production Plann.Cont. 1, 147-157.
[12] Tannock TDT, 2003, A fuzzy control charting methods for individuals, International Journal of Production Research, 41(5), 1017-1032.
[13] Taleb H, Limam M, 2002, On fuzzy and probabilistic control charts, International Journal of Production Research, 40(12), 2849-2863.
[14] Wang JH and Raz T, 1990, On the construction of control charts using linguistic variables, International Journal of Production Research, 28, 477-487.
[15] Wetherill GB, Brown DW, 1991, Statistical Process Control, Chapman and Hall, London.
[16] Zadeh LA, 1965, Fuzzy sets, Information and control, 8: 338-353.
[17] Zimmermann HJ, 1985, Fuzzy set theory and applications, Kluwer-Nijhoff Publishing.