Int. J. Appl. Math. Comput. Sci., 2015, Vol. 25, No. 2, 389–401
DOI: 10.1515/amcs-2015-0030

A FUZZY NONPARAMETRIC SHEWHART CHART BASED ON THE BOOTSTRAP APPROACH

DABUXILATU WANG [a], OLGIERD HRYNIEWICZ [b,*]

[a] Department of Statistics, School of Economics and Statistics
Guangzhou University, 230 Waihuanxi Road, Guangzhou, 510006, PR China
e-mail: wangdabu@gzhu.edu.cn

[b] Systems Research Institute
Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland
e-mail: hryniewi@ibspan.waw.pl

In this paper, we consider a nonparametric Shewhart chart for fuzzy data. We utilize the fuzzy data without transforming them into a real-valued scalar (a representative value). Usually fuzzy data (described by fuzzy random variables) do not have a distributional model available, and also the size of the fuzzy sample data is small. Based on the bootstrap methodology, we design a nonparametric Shewhart control chart in the space of fuzzy random variables equipped with some L_2 metric, in which a novel approach for generating the control limits is proposed. The control limits are determined by the necessity index of strict dominance combined with the bootstrap quantile of the test statistic. An in-control bootstrap ARL of the proposed chart is also considered.

Keywords: Shewhart control chart, fuzzy data, bootstrap, average run length.

1. Introduction

Statistical process control (SPC) is very important in that it is proven to bring processes into control and maintain the control condition (cf. Wetherill and Brown, 1991). Control charts are principle tools that have been designed and applied for the purpose of SPC, and Shewhart control charts are the most popular and widely applied SPC tools (cf. Nelson, 1985; Wetherill and Brown, 1991).

Typically, a control chart consists of a centre line (CL) and two control lines, referred to as the upper control limit (UCL) and the lower control limit (LCL). The centre line represents an estimate of the process level, while the two control limits denote the boundaries of normal variability, and are specified in such a way that the majority of the observations lie within their bounded range when the process is under control. Samples drawn from the process are plotted as points on the control chart. The control charts are constructed concurrently with the statistical hypotheses testing process. Essentially, the control chart tests the hypothesis that the process remains in the state of statistical control. Accordingly, a plotted data point falling within the control limits confirms the hypothesis of statistical control, while a point falling outside of the control limits indicates a rejection of this hypothesis (cf. Wetherill and Brown, 1991).

It is well known that control charts are based on data representing one or several quality-related characteristics of the product or service. If these characteristics are measurable and represented by real-valued numbers or vectors of numbers, variable control charts are used. If the quality-related characteristics cannot be easily represented in a numerical form, attribute control charts such as the p-chart or c-chart are useful (cf. Wetherill and Brown, 1991). However, when SPC is applied in areas such as economic quality control or acceptance samplings for attributes with data based on imprecise observations and measurements, the key process characteristics sometimes may be much more complicated. For instance, sample data collected from the evaluation of the color-intensity of produced pictures or screens are affected by imprecision, and can be described using interval-valued or vague data. Also the vague

*Corresponding author

data may come from judgments reflecting humans' partial knowledge or subjectivity while evaluating categories or attributes of inspected items, and such judgments may be expressed in some verbal form which cannot be appropriately expressed on a numerical scale. For instance, when the process quality characteristics are sense based attributes such as appearance, hardness, softness, color, taste, style, etc., results of observations are usually not numerical, but may be expressed by the linguistic terms like "very good", "good", "medium", "poor", etc. Control charts for attributes that are designed for the control of such processes, when the conventional binary classification into conforming and nonconforming is applied, might not be appropriate as the quality of a product does not change abruptly from perfect to worthless, and there might be a number of intermediate levels, e.g., such as "perfect", "good", "medium", "poor", "bad". There also exists the concept of user based quality which emphasizes that for customers the quality of a product is its suitability, rather than its conformity to certain standards (Cen, 1996). Customers' appraisal on quality is frequently expressed in linguistic terms involving vague attributes (Cen, 1996).

Therefore, traditional control charts, based on precise data to monitor processes, have to be expanded in order to possibly carry out process monitoring tasks in imprecise or fuzzy environments. The extended attribute control charts such as χ^2-control charts for multi-label categories or control charts for grouped data may be suitable for this purpose, but only if the vague data could be expressed approximately in distinct categorical form and the distributions of the underlying process variables are known (Woodall and Tucker, 1997). However, the uncertainty of the vague data is essentially non-statistical in nature so that the conditions mentioned above are usually hardly satisfied.

Fuzzy set theory and possibility theory provide useful tools for dealing with imprecise data (Zadeh, 1965; 1975). The applicability of fuzzy sets (Zadeh, 1965) for the description of quality has been explained in many papers, e.g., by Cen (1996), who proposed to control the fuzzy suitability quality derived from imprecise opinions of end-users.

There have been some publications dedicated to the design of control charts with linguistic or fuzzy data. In the case of monitoring unique fuzzy quality characteristics, Wang and Raz (1990) as well as Raz and Wang (1990) were the first authors who proposed a control chart for linguistic data (fuzzy data). They pointed out that linguistic data can provide more information than the binary classification used in control charts by attributes. They proposed representative values control charts with both the probability rule and the membership function rule, for which the linguistic data (fuzzy data) are transformed into scalars referred to as representative values of the fuzzy data, and these representative values are plotted on an ordinary control chart. In their paper four kinds of transformation formula have been proposed, i.e., the fuzzy mode, fuzzy midrange, fuzzy median and fuzzy average.

Kanagawa *et al*. (1993) proposed another representative values chart by using the barycenter of the fuzzy data, in which the required probability density function needs to be estimated using the Gram–Charlier series. This method may be used not only for monitoring the fuzzy process mean, but also for monitoring the process variability. The main difficulty of this approach is the determination of the unknown density function. Taleb and Limam (2002) discussed various procedures of the construction of control charts for linguistic data using representative values of Wang and Raz, based on fuzzy sets and probability models. They compared fuzzy and probabilistic approaches using the concept of the average run length and real-life data. The representative value, i.e., α-level *fuzzy midrange* method was also employed by Senturk and Erginel (2009), who designed $\widetilde{\overline{X}} - \widetilde{R}$ and $\widetilde{\overline{X}} - \widetilde{S}$ control charts for fuzzy data, which heavily rely on the properties of the normal distribution. In the case of monitoring multiple fuzzy quality characteristics, recently Taleb (2009) proposed a multivariate control chart for multivariate attribute processes based on representative values charting methods of Wang and Raz, and presented some applications of the multivariate attribute control chart in decorated porcelain production. This work is an excellent example of successful application of fuzzy control charts in a real production line.

The methodology of fuzzy statistical tests proposed by Kruse and Meyer (1987) was first applied for the construction of the Shewhart control chart and the EWMA (exponential weighted moving average) control chart with fuzzy data by Höppner and Wolff (1995) as well as Höppner (1994). Kruse and Meyer (1987) used the concept of a fuzzy random variable first proposed by Féron (1976) and further developed by Kwakernaak (1978; 1979), according to which a fuzzy random variable is a fuzzy perception of an original crisp random variable with a known distributional model. Grzegorzewski and Hryniewicz (2000) were the first authors who presented a Shewhart control chart with fuzzy random variable of Féron–Kwakernaak–Kruse and Meyer for which the degree of fuzziness was taken into account using the necessity index of strict dominance (NSD) for the design and operation purposes.

Gülbay and Kahraman (2006; 2007) not only explained why we require fuzzy control charts, but also carefully discussed the charts with the fuzzy random variable of the Féron–Kwakernaak–Kruse and Meyer model relying on the normal distribution with respect to the *fuzzy mode, fuzzy midrange*, and *fuzzy median*.

They took under consideration the α levels of fuzzy sets, showing that the inspection will becomes tighter as the α-level is set to higher values. Also without any defuzzification they propose a direct fuzzy approach for constructing a fuzzy c-chart, in which they calculate the percentage of the area under the membership function which shows that the fuzzy sample statistic remains inside the fuzzy control limits. In their approach a decision on whether the process is in control is made according to preferences of operators. This direct fuzzy charting is a completely novel reasonable method in the area of SPC. However, it is somewhat complicated for practical applications due to the computations of the area under the membership function which depends on the shape of this membership function and the selected α-level.

Faraz and Shapiro (2010) proposed a Shewhart chart for trapezoidal fuzzy data using a more general definition of the fuzzy random variable proposed by Puri and Ralescu (1986), in which under the given significance level the fuzzy in-control region (FIR) is first determined, and then a proper fuzzy inclusion operator is selected in order to determine the degree to which fuzzy sample groups are excluded from the FIR. This work provided another direct fuzzy approach for constructing control charts in a fuzzy environment, and here the normal distribution of the "normally distributed" fuzzy random variables is indeed concerned with the underlying process variable according to Gil *et al*. (2006). Shu and Wu (2011) presented a fuzzy $\overline{X}$ chart and R chart using an expanded fuzzy dominance approach based on the resolution identity of a fuzzy number.

It should be pointed out that most of the aforementioned works considered the Shewhart chart with representative values of fuzzy data, and only few works considered the Shewhart chart, c-chart and EWMA chart for fuzzy data for which the representative values methods have not been used, and in nearly all of them it was assumed that, as Faraz and Shapiro (2010) pointed out, the underlying process variable was normally distributed.

A representative value of fuzzy data may result in losing important information included in original data and the effectiveness of the representative value charts is also reduced (Gülbay and Kahraman, 2007; Faraz and Shapiro, 2010), so the direct fuzzy way of establishing control charts with fuzzy data is expected to be developed and improved. In most of the existing control charts with fuzzy data established under a fuzzy direct way (cf. Gülbay and Kahraman, 2007; 2006; Shu and Wu, 2011; Faraz and Shapiro, 2010) the underlying processes variables were assumed to be normally distributed, and there is no report considering the case where the probability distribution of an underlying process variable may be non-existing or unknown. We may recall the work of Cheng (2005), who proposed a fuzzy regression method for generating fuzzy data based on experts' score from evaluating product quality, and constructed a control chart using exclusively the fuzzy approach based on possibility theory, where no probability distribution models were considered for the underlying process variable, and therefore some randomness probably implied in the data considered was ignored. In this paper, we will establish a sort of distribution-free Shewhart control chart for the case where the probability distribution of the underlying quality variable is unknown or does not exist.

The rest of the article is organized as follows. In Section 2, some preliminary knowledge on fuzzy random variables and related concepts such as the L_2-distance between two fuzzy sets (also for fuzzy numbers), Aumann expectation, Fréchet variance and covariance, support function of a bounded convex set, LR-fuzzy random variable, etc. are presented. In Section 3, a novel, and completely distribution-free, Shewhart mean chart is proposed for fuzzy random variables using a bootstrap approach. In Section 4, some simulation examples are presented for explaining the performance of the proposed chart. Finally, conclusions are formulated.

2. Some statistics based on fuzzy data

2.1. Fuzzy sets on $\mathbb{R}^d$. A fuzzy set $\tilde{u}$ of $\mathbb{R}^d$ is equivalent to its membership function $\tilde{u} : \mathbb{R}^d \to [0,1]$, where the number $\tilde{u}(x)$ represents the degree of membership that x belongs to $\tilde{u}$. By $F(\mathbb{R}^d)$ we denote the collection of all normal, convex and compact fuzzy sets on $\mathbb{R}^d$, i.e., for $\tilde{u} \in F(\mathbb{R}^d)$, (i) there exists $x_0 \in \mathbb{R}^d$ such that $\tilde{u}(x_0) = 1$, (ii) the $\alpha-$cut of $\tilde{u}$, $\tilde{u}_\alpha := \{x \in \mathbb{R}^d : \tilde{u}(x) \geq \alpha\}$, $\alpha \in (0,1]$, is a convex and compact set of $\mathbb{R}^d$, (iii) $\tilde{u}_0 := cl\{x \in \mathbb{R}^d : \tilde{u}(x) > 0\}$, the support of $\tilde{u}$, is compact.

Zadeh's extension principle (cf. Zadeh, 1975) allows us to perform addition and scalar multiplication on $F(\mathbb{R}^d)$:

$$(\tilde{u} \oplus \tilde{v})(x) = \sup_{s+t=x} \min(\tilde{u}(s), \tilde{v}(t)), \quad x \in \mathbb{R}^d,$$

$$(a \odot \tilde{u})(x) = \begin{cases} \tilde{u}(\frac{x}{a}), & a \neq 0, \\ 0, & a = 0, \end{cases}$$

and, for any $a, b \in \mathbb{R}$

$$(ab) \odot \tilde{u} = a \odot (b \odot \tilde{u}),$$

$$a \odot (\tilde{u} \oplus \tilde{v}) = (a \odot \tilde{u}) \oplus (a \odot \tilde{v}).$$

But there holds only for $ab \geq 0, a, b \in \mathbb{R}$, that

$$(a + b) \odot \tilde{u} = (a \odot \tilde{u}) \oplus (b \odot \tilde{u}).$$

This indicates that $(F(\mathbb{R}^d), \oplus, \odot)$ is not a linear space. With Minkowski's sets operation, we have

$$(\tilde{u} \oplus \tilde{v})_\alpha = \tilde{u}_\alpha \oplus \tilde{v}_\alpha, \quad \alpha \in (0,1],$$

$$(a \odot \tilde{u})_\alpha = a \odot \tilde{u}_\alpha, \quad \alpha \in (0,1].$$

A support function of $\tilde{u} \in F(\mathbb{R}^d)$ is defined as

$$S_{\tilde{u}_\alpha}(x) = \begin{cases} \sup\limits_{t \in \tilde{u}_\alpha} \{x \cdot t\}, & \alpha \in (0,1], \\ 0, & \alpha = 0. \end{cases}$$

with $x \in S^{d-1} = \{x : \|x\| = 1\}$, where $\cdot$ denotes the inner product in the Euclidean space $\mathbb{R}^d$. For $\tilde{u}, \tilde{v} \in F(\mathbb{R}^d)$ and $a \in \mathbb{R}$, we have

$$S_{\tilde{u} \oplus \tilde{v}} = S_{\tilde{u}} + S_{\tilde{v}},$$

$$S_{a \odot \tilde{u}}(x) = \begin{cases} aS_{\tilde{u}}(x), & a > 0, \\ -aS_{\tilde{u}}(-x), & a < 0. \end{cases}$$

Thus, we get

$$\begin{aligned} &S_{((a \odot \tilde{u}) \oplus (b \odot \tilde{v}))_\alpha}(x) \\ &\quad = \begin{cases} (aS_{\tilde{u}_\alpha} + bS_{\tilde{v}_\alpha})(x), & a, b > 0, \\ -(aS_{\tilde{u}_\alpha} + bS_{\tilde{v}_\alpha})(-x), & a, b < 0, \end{cases} \end{aligned}$$

where $\alpha \in [0,1]$. The L_2-distance between $\tilde{u}, \tilde{v}$ and their scalar product are defined by (cf. Näther, 2000)

$$\begin{aligned} &\delta_2(\tilde{u}, \tilde{v}) \\ &:= \Big(\mathrm{d} \int_0^1 \int_{S^{d-1}} (S_{\tilde{u}_\alpha}(x) - S_{\tilde{v}_\alpha}(x))^2 \, \mu(\mathrm{d}x) \, \mathrm{d}\alpha \Big)^{1/2}, \end{aligned} \tag{1}$$

where μ is a normalized Lebesgue measure with $\mu(S^{d-1}) = 1$.

2.2. Fuzzy random variables of Puri–Ralescu. The idea of fuzzy random variables is inspired by the attempt to treat and model two different types of uncertainty, i.e., randomness and fuzziness, simultaneously. Let $(\Omega, \mathcal{A}, P)$ be a complete probability space. The mapping $\tilde{X} : \Omega \to F(\mathbb{R}^d)$ is said to be a fuzzy random variable (f.r.v.) if $\tilde{X}$ is $\mathcal{A} - \mathcal{B}$ measurable, where $\mathcal{B}$ is a σ-algebra induced by $\tilde{X}$ associated with δ_2.

Given a f.r.v. $\tilde{X}$, $S_{\tilde{X}_\alpha}$ is a random element and $E(S_{\tilde{X}_\alpha}) = S_{E(\tilde{X}_\alpha)}$ (cf. Näther, 2006; Gil *et al.*, 2006) if the expectation $E(\tilde{X}_\alpha)$ exists, where $E(\tilde{X}_\alpha)$ is an Aumann expectation of $(\tilde{X}_\alpha), \alpha \in [0,1]$ (cf. Puri and Ralescu, 1986; Gil *et al.*, 2006).

In the sequel, we assume that any f.r.v. $\tilde{X}$ fulfills the inequality

$$\mathrm{E}(\|\tilde{X}\|) := \mathrm{E}(\delta_2^2(\tilde{X}, \{0\})) < +\infty.$$

In contrast to the Féron–Kwakernaak–Kruse and Meyer model of the f.r.v., where the variance is defined as a simple fuzzy perception of the ordinary crisp variance, the measures of variability in the Puri–Ralescu model can be defined in various ways (see the work of Couso *et al.* (2007) for more information). In the most popular approach this variability is measured by the Fréchet variance of $\tilde{X}$ defined by Körner (1997) as

$$\begin{aligned} \mathrm{Var}(\tilde{X}) &:= \mathrm{E}(\delta_2^2(\tilde{X}, \mathrm{E}(\tilde{X}))) \\ &= \mathrm{d} \int_0^1 \int_{S^{d-1}} \mathrm{Var}(S_{\tilde{X}_\alpha}(x)) \, \mu(\mathrm{d}x) \, \mathrm{d}\alpha. \end{aligned}$$

Similarly, the covariance of two f.r.v.'s $\tilde{X}, \tilde{Y}$ is defined by

$$\begin{aligned} &\mathrm{Cov}(\tilde{X}, \tilde{Y}) \\ &\quad := \mathrm{d} \int_0^1 \int_{S^{d-1}} \mathrm{Cov}(S_{\tilde{X}_\alpha}(x), S_{\tilde{Y}_\alpha}(x)) \, \mu(\mathrm{d}x) \, \mathrm{d}\alpha. \end{aligned}$$

Note that

$$\begin{aligned} &\mathrm{Cov}((a \odot \tilde{X}) \oplus (b \odot \tilde{Y}), c \odot \tilde{Z}) \\ &\quad = ac\mathrm{Cov}(\tilde{X}, \tilde{Z}) + bc\mathrm{Cov}(\tilde{Y}, \tilde{Z}) \end{aligned}$$

holds only for $ac \geq 0, bc \geq 0, a, b, c \in \mathbb{R}$.

The independence of f.r.v.'s can be followed by the independence of the random elements, which is already defined (cf. Näther, 2000). Thus, obviously, if f.r.v.'s $\tilde{X}$ and $\tilde{Y}$ are independent, then $\mathrm{Cov}(\tilde{X}, \tilde{Y}) = 0$. However, if $\mathrm{Cov}(\tilde{X}, \tilde{Y}) \neq 0$, then they will not be independent. There is some sense of dependence between them; however, how to measure the dependencies between f.r.v.'s is still an open problem.

2.3. Fuzzy number and related statistics. When $d = 1$, a fuzzy set $\tilde{u}$ on $\mathbb{R}$ is said to be a fuzzy number. By $\mathcal{F}(\mathbb{R})$ we denote the set of all fuzzy numbers. Note that, for $\tilde{u}, \tilde{v} \in \mathcal{F}(\mathbb{R})$, we have

$$\begin{aligned} &2\delta_2^2(\tilde{u}, \tilde{v}) \\ &\quad = \int_0^1 \big[(S_{\tilde{u}_\alpha}(-1) - S_{\tilde{v}_\alpha}(-1))^2 \\ &\qquad + (S_{\tilde{u}_\alpha}(1) - S_{\tilde{v}_\alpha}(1))^2 \big] \, \mathrm{d}\alpha \\ &\quad = \int_0^1 \big[(\tilde{u}_\alpha^+ - \tilde{v}_\alpha^+)^2 + (\tilde{u}_\alpha^- - \tilde{v}_\alpha^-)^2 \big] \, \mathrm{d}\alpha \\ &\quad = d_*^2(\tilde{u}, \tilde{v}), \end{aligned} \tag{2}$$

where the intervals $[\tilde{u}_\alpha^-, \tilde{u}_\alpha^+], [\tilde{v}_\alpha^-, \tilde{v}_\alpha^+]$ denote the α level sets of $\tilde{u}, \tilde{v}$, respectively, d_* denotes the distance between two fuzzy numbers proposed by Feng *et al.* (2001).

The following parametric class of fuzzy numbers, the so-called LR-fuzzy numbers, are often used in applications:

$$\tilde{u}(x) = \begin{cases} L\Big(\dfrac{m - x}{l}\Big), & x \leq m, \\ R\Big(\dfrac{x - m}{r}\Big), & x > m. \end{cases} \tag{3}$$

Here $L : \mathbb{R}^+ \to [0,1]$ and $R : \mathbb{R}^+ \to [0,1]$ are given left-continuous and non-increasing functions with $L(0) = R(0) = 1$. L and R are respectively called left and right shape functions, m is the central point of $\tilde{u}$, and $l > 0$, $r > 0$ are respectively the left and right spreads of $\tilde{u}$. An LR-fuzzy number is abbreviated by $\tilde{u} = (m,l,r)_{LR}$. Particularly, $(m,0,0)_{LR} := m$. It has been proven that LR-fuzzy numbers possess some nice properties for operations:

$$\begin{aligned}(m_1,l_1,r_1)_{LR} &\oplus (m_2,l_2,r_2)_{LR}\\ &= (m_1+m_2, l_1+l_2, r_1+r_2)_{LR},\end{aligned}$$

$$a \odot (m,l,r)_{LR} = \begin{cases} (am, al, ar)_{LR}, & a > 0,\\ (am, -ar, -al)_{RL}, & a < 0,\\ 0, & a = 0,\end{cases}$$

$$(m_1,l_1,r_1)_{LR} \ominus m_2 = (m_1 - m_2, l_1, r_1)_{LR}.$$

The last equality can be understood as meaning that the fuzzy number $(m_1,l_1,r_1)_{LR}$ has a shift from m_1 to m_2.

Let

$$L^{(-1)}(\alpha) := \sup\{x \in R | L(x) \geq \alpha\},$$

$$R^{(-1)}(\alpha) := \inf\{x \in R | R(x) \geq \alpha\}.$$

Then for $\alpha \in [0,1]$

$$\tilde{u} = (m,l,r)_{LR},$$

$$\tilde{u}_\alpha = \left[m - lL^{(-1)}(\alpha), \quad m + rR^{(-1)}(\alpha)\right].$$

Körner (2000) defined a LR-*f.r.v.* on the probability space $(\Omega, \mathcal{A}, P)$ as a measurable mapping $\tilde{X} : \Omega \to \mathcal{F}_{LR}(\mathbb{R})$, $\tilde{X}(\omega) = (m(\omega), l(\omega), r(\omega))_{LR}$, $\omega \in \Omega$. In short, we write it $\tilde{X} = (m,l,r)_{LR}$, where m, l, r are three real-valued random variables with $P\{l \geq 0\} = P\{r \geq 0\} = 1$. In a fuzzy observation on objects of interest, the outcomes can be viewed as LR-fuzzy data under a proper assumption, i.e., the data are viewed as realizations of LR-f.r.v..

Note that, in fact, the three random variables m, l, r are usually dependent, as they integrate a fuzzy number $(m,l,r)_{LR}$ together with $L(x), R(x)$, which is affected by the common factor of the data fuzziness, e.g., if the data are more fuzzy, then l, r become larger, and if the data are more crisp, then l, r become smaller, while m is the center location of the fuzzy data.

For an LR-f.r.v. $\tilde{X}$, its expectation of the Aumann type is as follows:

$$\mathrm{E}(\tilde{X}) = (\mathrm{E}(m), \mathrm{E}(l), \mathrm{E}(r))_{LR},$$

and the Fréchet variance is

$$\begin{aligned}\mathrm{Var}(\tilde{X}) = \frac{1}{2}\int_0^1 &(\mathrm{Var}(m - lL^{(-1)}(\alpha))\\ &+ \mathrm{Var}(m + rR^{(-1)}(\alpha)))\,\mathrm{d}\alpha.\end{aligned}$$

Let $\tilde{X}_1, \ldots, \tilde{X}_n$ be a sample of size n from $\tilde{X} = (m,l,r)_{LR}$ under independent observations. Then the sample mean $\overline{\tilde{X}}$ and the sample variance S_n^2 are defined by

$$\overline{\tilde{X}} := \frac{1}{n}\sum_{i=1}^n \tilde{X}_i = (\overline{m}, \overline{l}, \overline{r})_{LR}, \tag{4}$$

$$S_n^2 := \frac{1}{(n-1)}\sum_{i=1}^n \delta_2^2(\tilde{X}_i, \overline{\tilde{X}}), \tag{5}$$

i.e.,

$$\begin{aligned} S_n^2 &= \frac{1}{2(n-1)}\sum_{i=1}^n \int_0^1 \Big[((m_i - \overline{m}) + (\overline{l} - l_i)L^{(-1)}(\alpha))^2\\ &\quad + ((m_i - \overline{m}) + (r_i - \overline{r})R^{(-1)}(\alpha))^2\Big]\,\mathrm{d}\alpha \end{aligned} \tag{6}$$

and

$$\overline{m} = \frac{1}{n}\sum_{i=1}^n m_i,$$

$$\overline{l} = \frac{1}{n}\sum_{i=1}^n l_i,$$

$$\overline{r} = \frac{1}{n}\sum_{i=1}^n r_i.$$

Lemma 1. (Näther, 2000) *Let $\tilde{X}_1, \ldots, \tilde{X}_n$ be a fuzzy random sample from an f.r.v. $\tilde{X}$ defined on probability space $(\Omega, \mathcal{A}, P)$ and with values in $F(\mathbb{R}^d)$. Then*

$$\delta_2(\overline{\tilde{X}}\mathrm{E}(\tilde{X})) \to_P 0 \quad \text{as } n \to \infty, \tag{7}$$

$$\begin{aligned}&\mathrm{E}(S_n^2) = \mathrm{Var}(\tilde{X}),\\ &S_n^2 \to_P \mathrm{Var}(\tilde{X}) \quad \text{as } n \to \infty,\end{aligned} \tag{8}$$

where $\mathrm{E}(\tilde{X}))$ and $\mathrm{Var}(\tilde{X})$ are the Aumann expectation and the Fréchet variance of the f.r.v. $\tilde{X}$, respectively, and

$$S_n^2 = \frac{1}{(n-1)}\sum_{i=1}^n \delta_2^2(\tilde{X}_i, \overline{\tilde{X}})$$

is the Fréchet sample variance.

In making statistical decisions, there is a need for a comparison of fuzzy numbers (see the work of Grzegorzewski and Hryniewicz (2000) for an application in statistical quality control). For this purpose, we may employ the necessity index of strict dominance (NSD) introduced by Dubois and Prade (1983). Let $\tilde{u}$ and $\tilde{v}$ be two fuzzy numbers. The necessity of strict dominance of $\tilde{u}$ over $\tilde{v}$ is calculated from the formula

$$\mathrm{NSD} = \mathrm{Ness}(\tilde{u} > \tilde{v}) = 1 - \sup_{x,y:x \leq y}[\min\{\tilde{u}(x), \tilde{v}(y)\}]. \tag{9}$$

If there exists some constant $c_0 \in [0,1]$ such that $\mathrm{Ness}(\tilde{u} > \tilde{v}) > c_0$, then we may say that $\tilde{u}$ dominates $\tilde{v}$ to the degree of c_0. The calculation of the NSD index is simplified when the compared numbers are of the LR type, defined by (3).

Let $\tilde{u} = (m_u, l_u, r_u)_{LR}$ and $\tilde{v} = (m_v, l_v, r_v)_{L'R'}$ be two fuzzy numbers. Note that the shape curves of the functions L and L' (R and R') may be different. Let z_0 be the solution of the equation

$$L\left(\frac{m_u - z_0}{l_u}\right) = R'\left(\frac{z_0 - m_v}{l_v}\right).$$

Then, we have

$$\begin{aligned}&\mathrm{Ness}(\tilde{u} > \tilde{v})\\ &\quad = \begin{cases} 0, & m_u \le m_v,\\ 1 - L\left(\dfrac{m_u - z_0}{l_u}\right), & m_v \le z_0 \le m_u, \end{cases}\end{aligned} \tag{10}$$

Note that we can also define a necessity index of inequality (NII) of $\tilde{u}$ with $\tilde{v}$ as

$$\begin{aligned}\mathrm{NII} &= \mathrm{Ness}(\tilde{u} \neq \tilde{v})\\ &= 1 - \sup_{x,y:x=y} [\min\{\tilde{u}(x), \tilde{v}(y)\}].\end{aligned} \tag{11}$$

One possible application of the NII is that it could be useful for the so-called two-sided fuzzy hypotheses testing based on the necessity measure.

3. Shewhart chart for fuzzy data

Let us consider the conventional Shewhart control chart (cf. Nelson, 1985) for monitoring the process mean. Under the assumption that the process variable obeys normal distribution $N(\mu_0, \sigma^2)$, the well-known Shewhart chart is given by the following three lines:

$$\begin{aligned}\mathrm{UCL} &= \mu_0 + z_{1-\delta/2}\frac{\sigma}{\sqrt{m}},\\ \mathrm{CL} &= \mu_0,\\ \mathrm{LCL} &= \mu_0 - z_{1-\delta/2}\frac{\sigma}{\sqrt{m}},\end{aligned}$$

where m is the sample size and $z_{1-\delta/2}$ is the $1-\delta/2$ quantile of the standard normal distribution. In practice, the parameters μ and σ of the model need to be estimated based on the so-called Phase I samples, and therefore the control limits of the Shewhart chart are determined not only by the sample size but also the number of Phase I samples.

However, sometimes the distributional models of the process variables are unknown, especially when the observations are vague. In such cases one may employ non-parametric statistical methods to estimate the process model and the parameters concerned, and then design a non-parametric Shewhart control chart (see, e.g., Liu and Tang, 1996). As mentioned in the previous sections, we consider the case where the underlying process variable is an f.r.v. which is distribution-free. Thus, for such a case the construction of a Shewhart chart using a nonparametric method has to be proposed. Recalling that the previous charts with fuzzy data (see the Introduction) are based on the assumption that the examined underlying process quality variable is normally distributed, i.e., though the observational linguistic or score results indirectly appear as fuzzy data (human perception), behind them there exists a normally distributed random variable which can be taken to be a model for statistical testing and inference. If now the underlying variable becomes completely fuzzy, then there may be no distributional models behind the observation fuzzy data, so that no models can be employed. For example, suitability quality (Cen, 1996) could be considered a fuzzy quality because it is proposed based on fuzzy information (or linguistic information) from a collection of all opinions expressed linguistically or scores provided by end-users. A sort of fuzzy regression method for generating an LR-fuzzy number from the expert's opinions has been proposed by Cheng (2005).

When fuzzy random data are generated according to the fuzzy Puri and Ralescu model, there exist serious problems with the construction of statistical tests, since such data are usually distribution free. These problems can be overcome well by applying statistical bootstrap techniques, as proposed by Efron and Tibshirani (1993), Montenegro *et al.* (2004) or Liu and Tang (1996). In the case when the underlying process variable is a (crisp) random variable, bootstrap sampling consists in taking (with replacement) bootstrap random samples $\{X_1^\star, X_2^\star, \ldots, X_m^\star\}$ of size m from a bootstrap population $\{X_1, X_2, \ldots, X_M\}$ consisting of M *independent* observations. Suppose that the bootstrap population is generated from a probability distribution F with mean μ and finite variance, and its empirical distribution is denoted by F_M. Now, let $\bar{X}_m^\star$ be random variable that describes the mean value of the bootstrap sample, and $\bar{X}_m$ be the random mean of the bootstrap population. If a sufficiently large number B of bootstrap samples is taken, then we have (Liu and Tang, 1996) that, almost surely,

$$\begin{aligned}&P(\sqrt{m}(\bar{X}_m^\star - \bar{X}_M) \le x | F_M)\\ &\qquad = P(\sqrt{m}(\bar{X}_m - \mu) \le x | F).\end{aligned} \tag{12}$$

The fundamental bootstrap equation (12) allows using the empirical distribution of $\sqrt{m}(\bar{X}_m^\star - \bar{X}_M)$ (a histogram of resulting B terms of $\sqrt{m}(\bar{X}_m^\star - \bar{X}_M)$) as an approximation of the unknown distribution of $\sqrt{m}(\bar{X}_m - \mu)$. Thus, this bootstrap distribution can be used for the construction of the Shewhart $\bar{X}$ control chart. To this end, for the given significance level α, we calculate lower and upper quantiles, $\tau_{\alpha/2}$ and $\tau_{1-\alpha/2}$, of the bootstrap histogram, such that

$$P(\sqrt{m}(\bar{X}_m^\star - \bar{X}_M) \leq \tau_{\alpha/2}|F_M) = \alpha/2 \quad \text{a.s.} \tag{13}$$

and

$$P(\sqrt{m}(\bar{X}_m^\star - \bar{X}_M) \leq \tau_{1-\alpha/2}|F_M) = 1-\alpha/2 \quad \text{a.s.} \tag{14}$$

Hence, the lower and upper limits of the bootstrap-based Shewhart $\bar{X}$ control chart are computed using the following formula (Liu and Tang, 1996):

$$\text{LCL} = \bar{X}_M + \frac{\tau_{\alpha/2}}{\sqrt{m}}, \tag{15}$$

$$\text{UCL} = \bar{X}_M + \frac{\tau_{1-\alpha/2}}{\sqrt{m}}. \tag{16}$$

Following the bootstrap-based (crisp) $\bar{X}$ chart described above, we now consider the construction of its fuzzy counterpart. In the case when the underlying process variable is a distribution-free f.r.v. $\tilde{X}$, and if a sequence of independent observations $\tilde{X}_1, \ldots, \tilde{X}_n$ on $\tilde{X}$ is generated, then according to Körner (2000) it follows that the random variable $\sqrt{n}\delta_2(\bar{\tilde{X}}_n, E(\tilde{X}))$ converges in distribution, as n tends to ∞, to a Gaussian random element. This conclusion can be used for the approximation of a sampling distribution model of the distance variable. In order to design the chart, we assume that at Phase I we can determine the "in-control" target value of the process mean in the following way. Under the "in-control" state inspectors draw independently n samples of size m from the observed fuzzy process. Each sample element is a term evaluated by score or linguistic form that can be transformed into fuzzy numbers with the method of Cheng (2005). Thus, all available observations are now described by fuzzy numbers $\tilde{x}_{i1}, \ldots, \tilde{x}_{im}$, $i = 1, 2, \ldots, n$. Naturally, we would take the target value as the fuzzy mean

$$\tilde{\mu}_0 := \overline{\overline{\tilde{X}}} = \frac{1}{mn} \bigoplus_{i=1}^{n} \bigoplus_{j=1}^{m} \tilde{x}_{ij}. \tag{17}$$

Moreover, we can evaluate the variability of our fuzzy observations as the average value of n within-group sample variations

$$\overline{\overline{s}} := \frac{1}{n-1} \sum_{i=1}^{n} \sqrt{\frac{1}{m-1} \sum_{j=1}^{m} \delta_2^2(\tilde{x}_{ij}, \bar{\tilde{x}}_i)}, \tag{18}$$

where

$$\bar{\tilde{x}}_i := \frac{1}{m} \bigoplus_{j=1}^{m} \tilde{x}_{ij}$$

is the within-group mean of the i-th group, $i = 1, 2, \ldots, n$, and $\delta_2^2(\cdot)$ is the distance between two fuzzy numbers, calculated according to (2). Based on the mn fuzzy observations $\tilde{x}_{i1}, \ldots, \tilde{x}_{im}$, $i = 1, 2, \ldots, n$ from the process under the "in-control" state, we can form a bootstrap population made of n independent fuzzy sample means $\{\bar{\tilde{x}}_1, \cdots, \bar{\tilde{x}}_n\}$. Then from this population we take a bootstrap sample of B elements denoted by

$$\tilde{x}^{*^b} := \{\tilde{x}_1^{*^b}, \tilde{x}_2^{*^b}, \ldots, \tilde{x}_k^{*^b}\}, \quad b = 1, \ldots, B, \tag{19}$$

and for each b we denote the sample mean of each bootstrap element by

$$\overline{\tilde{x}^*}^b := \frac{1}{k} \bigoplus_{j=1}^{k} \tilde{x}_j^{*^b},$$

where $1 \leq k \leq n$. In the next step of our procedure we build the bootstrap distribution (bootstrap histogram) of *distances* between the fuzzy target value $\tilde{\mu}_0$ and the sample means of the bootstrap elements, defined as

$$u^{*^b} := \sqrt{k}\delta_2(\overline{\tilde{x}^*}^b, \tilde{\mu}_0), \quad b = 1, 2, \ldots, B. \tag{20}$$

According to Liu and Tang (1996) as well as Körner (2000) the distance variable u^{*^b} may be approximated by a Gaussian random element of a Hilbert space which follows the ω^2-distribution. Thus, we denote by $u_{\alpha/2}$ and $u_{1-\alpha/2}$, respectively, the two-sided empirical quantiles of the bootstrap distribution of u^{*^b}. The fuzzy bootstrap-based Shewhart control chart can be defined by analogy to (15)–(16) as follows:

$$\widetilde{\text{LCL}} = \tilde{\mu}_0 + \frac{u_{\alpha/2}}{\sqrt{k}}, \tag{21}$$

$$\widetilde{\text{UCL}} = \tilde{\mu}_0 + \frac{u_{1-\alpha/2}}{\sqrt{k}}. \tag{22}$$

The users who are accustomed to traditional description of the Shewhart control chart can rewrite (21) and (22) as

$$\widetilde{\text{LCL}} = \tilde{\mu}_0 + u'_{\alpha/2}\frac{\overline{\overline{s}}}{\sqrt{k}}, \tag{23}$$

$$\widetilde{\text{UCL}} = \tilde{\mu}_0 + u'_{1-\alpha/2}\frac{\overline{\overline{s}}}{\sqrt{k}}, \tag{24}$$

where $u'_{\alpha/2} = u_{\alpha/2}/\overline{\overline{s}}$ and $u'_{1-\alpha/2} = u_{1-\alpha/2}/\overline{\overline{s}}$.

Note that in the case considered the control limits are given as fuzzy numbers $\widetilde{\text{LCL}}, \widetilde{\text{UCL}}$, and the observed sample means $\bar{\tilde{x}}$ are also expressed as fuzzy numbers. Therefore, in order to make decisions we have to compare fuzzy observations ($\bar{\tilde{x}}$) with fuzzy control limits $\widetilde{\text{LCL}}, \widetilde{\text{UCL}}$. For this comparison we may use the methodology proposed by Grzegorzewski and Hryniewicz (2000), which is based on the NSD index defined by (9) as well as the NII defined by (11).

Let ξ be the required level of the necessity, such that $\text{Ness}(\bar{\tilde{x}} \neq \tilde{\mu}_0) \geq \xi$, and $\left[(\bar{\tilde{x}})_{1-\xi}^-, (\bar{\tilde{x}})_{1-\xi}^+\right]$ be the

$(1-\xi)$-level set of the fuzzy mean value $\tilde{\bar{x}}$ calculated from the sample taken from a process. The definition of a possibilistic confidence interval (cf. Kruse and Meyer, 1987; Hryniewicz, 2006) implies that for the assumed value ξ the monitored process should be regarded as being fully in the "in-control" state if the following requirement is fulfilled:

$$\big[(\tilde{\bar{x}})^-_{1-\xi}, (\tilde{\bar{x}})^+_{1-\xi}\big] \subset \big[\widetilde{\mathrm{LCL}}^-_{1-\xi}, \widetilde{\mathrm{UCL}}^+_{1-\xi}\big], \tag{25}$$

where

$$\widetilde{\mathrm{LCL}}^-_{1-\xi} = (\tilde{\mu}_0)^-_{1-\xi} + \frac{u_{\alpha/2}}{\sqrt{k}}, \tag{26}$$

$$\widetilde{\mathrm{UCL}}^+_{1-\xi} = (\tilde{\mu}_0)^+_{1-\xi} + \frac{u_{1-\alpha/2}}{\sqrt{k}}, \tag{27}$$

and $\big[(\tilde{\mu}_0)^-_{1-\xi}, (\tilde{\mu}_0)^+_{1-\xi}\big]$ is the $(1-\xi)$-level set of the fuzzy central area $\tilde{\mu}_0$.

Otherwise, the process may be considered either to be only partially in the "in-control" state or to be completely out of control. The degree β to which the process considered can be regarded as being in the "in-control" state can be evaluated from the following formula:

$$\beta := \frac{\big\|\big[(\tilde{\bar{x}})^-_{1-\xi}, (\tilde{\bar{x}})^+_{1-\xi}\big] \cap \big[\widetilde{\mathrm{LCL}}^-_{1-\xi}, \widetilde{\mathrm{UCL}}^+_{1-\xi}\big]\big\|}{\big\|\big[\widetilde{\mathrm{LCL}}^-_{1-\xi}, \widetilde{\mathrm{UCL}}^+_{1-\xi}\big]\big\|}, \tag{28}$$

where $\|A\|$ denotes the length of an interval A, and the case $\big[\widetilde{\mathrm{LCL}}^-_{1-\xi}, \widetilde{\mathrm{UCL}}^+_{1-\xi}\big] \subset \big[(\tilde{\bar{x}})^-_{1-\xi}, (\tilde{\bar{x}})^+_{1-\xi}\big]$ is excluded.

Remark 1.
(i) The decision procedure described above can be interpreted according to the possibilistic interpretation of fuzzy statistical tests proposed by Hryniewicz (2006). Let $\big[(\tilde{\mu}_0)^-_{1-\alpha}, (\tilde{\mu}_0)^+_{1-\alpha}\big]$ be the $(1-\alpha)$-level set of the fuzzy number $\tilde{\mu}_0$. Then the interval $[C^-, C^+]$ forms a confidence interval of level $1-\alpha$, where

$$\begin{aligned} C^- &:= (\tilde{\mu}_0)^-_{1-\alpha} + u_{\alpha/2}/\sqrt{k} \\ &= (\tilde{\mu}_0)^-_{1-\alpha} + u'_{\alpha/2}\frac{\bar{\bar{s}}}{\sqrt{k}}, \end{aligned} \tag{29}$$

$$\begin{aligned} C^+ &:= (\tilde{\mu}_0)^-_{1-\alpha} + u_{1-\alpha/2}/\sqrt{k} \\ &= (\tilde{\mu}_0)^-_{1-\alpha} + u'_{1-\alpha/2}\frac{\bar{\bar{s}}}{\sqrt{k}} \end{aligned} \tag{30}$$

and such an interval $[C^-, C^+]$ can be used for the construction of the test for the fuzzy hypothesis $H_0 : \tilde{\bar{X}} = \tilde{\mu}_0$. Following the way of reasoning introduced by Hryniewicz (2006), we can claim that if the $(1-\alpha)$-level set of the fuzzy number $\tilde{\mu}_0$ is included in the interval $[C^-, C^+]$ then $\mathrm{Ness}(\tilde{\bar{x}} \neq \tilde{\mu}_0) \geq \alpha$, which means that α can be the required necessity level, and C^-, C^+ can be the lower and upper control limits denoted by LCL, UCL, respectively, in the decision stage when using the fuzzy Shewhart control chart defined by (21) and (22) (or by (23) and (24)). Thus, the required necessity index ξ can be replaced by α in the decision rule described by the formulas (25)–(27).

(ii) From the possibilistic interpretation of fuzzy statistical tests proposed by Hryniewicz (2006), the level value α can be viewed as a significance level used in hypotheses tests, the required necessity level for fuzzy decision and the confidential level as well as a membership value for a fuzzy set based on the definition of the possibilistic confidence interval, though the meanings of these notions remain different. The reason may be that the level value α is always given by experts in decisions making.

(iii) The required necessity level ξ is usually given by experts and used only for the comparison of fuzzy numbers. However, the notion of the significance level plays an important role in statistical hypotheses tests. Sometimes we may be allowed to use the significance level only. Then it may be reasonable to determine the required necessity level value from the given significance level α by the bootstrap approach if we really need it. Such a necessity level value is denoted by ξ_α and obtained from the following bootstrap equality:

$$\frac{\#\{\overline{\tilde{x}^*}^b : \mathrm{Ness}(\overline{\tilde{x}^*}^b \neq \tilde{\mu}_0) \leq \xi_\alpha\}}{B} = \alpha, \tag{31}$$

where $\#\{\cdot\}$ denotes the number of elements in the set $\{\cdot\}$.

We can now briefly summarize the whole procedure proposed above.

Step 1. Set the target value (17) based on nm fuzzy numbers generated from the experts' evaluation.

Step 2. Set the bootstrap population made of n independent fuzzy sample means $(\tilde{\bar{x}}_1, \cdots, \tilde{\bar{x}}_n)$.

Step 3. Take the bootstrap sample (19), and for each of its elements calculate the value of the bootstrap statistic according to (20).

Step 4. For the assumed value of the probability of false alarm α find the lower (of order $\alpha/2$) and upper (of order $1-\alpha/2$) empirical quantiles of the bootstrap statistic (20).

Step 5. Set the lower control limit (29) and the upper control limit (30).

Step 6. Make the decision according to (25)–(27) with ξ replaced by α.

The efficiency of a control chart is usually measured by the average run length (ARL). For the bootstrap-based control chart we can formulate the following proposition.

Proposition 1. *For a given value $\alpha \in (0,1)$ of the probability of a false alarm and the aforementioned control limits of the chart, the fraction of inspections while the process remains in control is bounded from below by approximately $1-\alpha$, and the bootstrap based in-control average run length $ARL_0^{*^b}$ is smaller than $1/\alpha$, approximately.*

Proof. For the given value α of the significance level and the required necessity level ξ, we obtain the quantiles $u_{\alpha/2}$ and $u_{1-\alpha/2}$ based on the bootstrap procedure. Note that $u_{1-\alpha/2} > u_{\alpha/2} > 0$ since distance variable takes nonnegative values. From the definition of the distance δ_2 given by (2), we have that

$$\frac{(\overline{\tilde{X}}^-_{1-\xi} - (\tilde{\mu}_0)^-_{1-\xi})^2 + (\overline{\tilde{X}}^+_{1-\xi_\alpha} - (\tilde{\mu}_0)^+_{1-\xi_\alpha})^2}{2} \leq \delta_2^2(\overline{\tilde{X}}, \tilde{\mu}_0).$$

On the other hand,

$$\frac{1}{B}\#\Big\{\frac{u^2_{\alpha/2}}{k} \leq \delta_2^2(\overline{\tilde{x}^*}^b, \tilde{\mu}_0) \leq \frac{u^2_{1-\alpha/2}}{k}\Big\} = 1-\alpha,$$

and

$$\delta_2^2(\overline{\tilde{x}^*}^b, \tilde{\mu}_0) \leq \delta_2^2(\overline{\tilde{X}}, \tilde{\mu}_0).$$

Thus, both

$$\frac{(\overline{\tilde{X}}^-_{1-\xi} - (\tilde{\mu}_0)^-_{1-\xi})^2 + (\overline{\tilde{X}}^+_{1-\xi} - (\tilde{\mu}_0)^+_{1-\xi})^2}{2}$$

and $\delta_2^2(\overline{\tilde{x}^*}^b, \tilde{\mu}_0)$ are less than $\delta_2^2(\overline{\tilde{X}}, \tilde{\mu}_0)$, which means that there is no big difference between the former two values. ■

Set

$$B_1 := \Big\{\overline{\tilde{X}} : \frac{u_{\alpha/2}}{\sqrt{k}} \leq \overline{\tilde{X}}^-_{1-\xi} - (\tilde{\mu}_0)^-_{1-\xi} \leq \frac{u_{1-\alpha/2}}{\sqrt{k}}\Big\},$$

$$B_2 := \Big\{\overline{\tilde{X}} : \frac{u_{1-\alpha/2}}{\sqrt{k}} \leq \overline{\tilde{X}}^-_{1-\xi} - (\tilde{\mu}_0)^-_{1-\xi}\Big\},$$

$$C_1 := \Big\{\overline{\tilde{X}} : \frac{u_{\alpha/2}}{\sqrt{k}} \leq \overline{\tilde{X}}^+_{1-\xi} - (\tilde{\mu}_0)^+_{1-\xi} \leq \frac{u_{1-\alpha/2}}{\sqrt{k}}\Big\},$$

$$C_2 := \Big\{\overline{\tilde{X}} : \overline{\tilde{X}}^+_{1-\xi} - (\tilde{\mu}_0)^+_{1-\xi} \leq \frac{u_{\alpha/2}}{\sqrt{k}}\Big\}.$$

The fraction of inspections while the process remains in control will be

$$\begin{aligned}
f_0 &:= \frac{1}{B}\#\Big\{\overline{\tilde{X}} : [(\overline{\tilde{X}})^-_{1-\xi}, (\overline{\tilde{X}})^+_{1-\xi}] \\
&\quad \subset [\widetilde{LCL}^-_{1-\xi_\alpha}, \widetilde{UCL}^+_{1-\xi_\alpha}]\Big\} \\
&= \frac{1}{B}\#\Big\{\overline{\tilde{X}} : (\overline{\tilde{X}}^-_{1-\xi} - \tilde{\mu}_0)^-_{1-\xi} \geq \frac{u_{\alpha/2}}{\sqrt{k}}, \\
&\quad \overline{\tilde{X}}^+_{1-\xi} - (\tilde{\mu}_0)^+_{1-\xi} \leq \frac{u_{1-\alpha/2}}{\sqrt{k}}\Big\} \\
&= \frac{1}{B}\#\Big\{B_1C_1 \cup B_2C_1 \cup B_1C_2 \cup B_2C_2\Big\} \\
&\approx \frac{1}{B}\Big[\#\{u^2_{\alpha/2} \leq k\delta_2^2(\overline{\tilde{x}^*}^b, \tilde{\mu}_0) \leq u^2_{1-\alpha/2}\} \\
&\quad + \#\{2k\delta_2^2(\overline{\tilde{x}^*}^b, \tilde{\mu}_0) \geqslant u^2_{\alpha/2} + u^2_{1-\alpha/2}\} \\
&\quad + \#\{u^2_{\alpha/2} \leq 2k\delta_2^2(\overline{\tilde{x}^*}^b, \tilde{\mu}_0) \leq u^2_{1-\alpha/2} + u^2_{\alpha/2}\} \\
&\quad + \#\{u^2_{\alpha/2} + u^2_{1-\frac{\alpha}{5}} \geq 2k\delta_2^2(\overline{\tilde{x}^*}^b, \tilde{\mu}_0) > u^2_{1-\alpha/2}\}\Big] \\
&= (1-\alpha) + (1-p_1(\alpha)) + (p_1(\alpha) - p_2(\alpha)) \\
&\quad + (p_4(\alpha) - p_3(\alpha)) \\
&\geq 1-\alpha,
\end{aligned}$$

where the probabilities $p_1(\alpha), p_2(\alpha), p_3(\alpha), p_4(\alpha)$ satisfy

$$u_{p_1(\alpha)} = \frac{u^2_{\alpha/2} + u^2_{1-\alpha/2}}{2}, \qquad u_{p_2(\alpha)} = \frac{u^2_{\alpha/2}}{2},$$

$$u_{p_3(\alpha)} = \frac{u^2_{1-\alpha/2}}{2}, \qquad u_{p_4(\alpha)} = \frac{u^2_{\alpha/2} + u^2_{1-\alpha/5}}{2}.$$

In consequence, we have $f_0 \geq 1-\alpha$, approximately, so that

$$\mathrm{ARL}_0^{*^b} = \frac{1}{1-f_0} \leq \frac{1}{\alpha},$$

approximately.

4. Applications and comparisons

In this section, an application of the proposed bootstrap based fuzzy Shewhart control chart is considered. Let us consider an artificial example inspired by the real example described by Taleb (2009). In a porcelain decorating process control the color appearance condition of the porcelain is one of the monitored quality characteristics. Assume that at Phase I we obtained 8 groups of size 5 decorated porcelains whose color appearance conditions are evaluated by experts with LR-fuzzy data shown in Table 1, where the membership shape functions are taken to be $L(x) = R(x) = \max\{0, 1-x\}$, and for the sample fuzzy data $\tilde{x}_{ij} = (m_{ij}, l_{ij}, r_{ij})_{LR}, i = 1,\dots,8$ and $j = 1,\dots,5$, the within-group mean $\overline{\tilde{x}}_i = (\overline{m}_i, \overline{l}_i, \overline{r}_i)_{LR}$, we have

$$\begin{aligned}\delta_2^2(\tilde{x}_{ij}, \overline{\tilde{x}}_i) &= (m_{ij} - \overline{m}_i)^2 + \frac{1}{2}(m_{ij} - \overline{m}_i)(\overline{l}_i - l_{ij} + r_{ij} - \overline{r}_i) \\ &\quad + \frac{1}{6}((\overline{l}_i - l_{ij})^2 + (r_{ij} - \overline{r}_i)^2).\end{aligned}$$

Let the set of the sample means $\{\overline{\tilde{x}}_1, \overline{\tilde{x}}_2, \overline{\tilde{x}}_3, \overline{\tilde{x}}_4, \overline{\tilde{x}}_5, \overline{\tilde{x}}_6, \overline{\tilde{x}}_7, \overline{\tilde{x}}_8\}$ be the bootstrap population, from which we randomly take a bootstrap sample of size 8 with replacement, and repeat this procedure a large number B of times, where $B = 10000$. The above bootstrap samplings and computations were performed using SAS software.

Assume that the given significance level α and the required necessity level ξ take the same value, $\alpha = 0.084 = \xi$. From the bootstrap experiment we have the following bootstrap quantiles: $u_{0.042} = 0.1341, u_{0.958} = 1.9817$. Thus, the control limits (26)–(27) for the above in control 40 fuzzy data are as follows:

$$\widetilde{\mathrm{LCL}}^{-}_{0.916} = 6.46202, \quad \widetilde{\mathrm{UCL}}^{+}_{0.916} = 8.1586.$$

By Proposition 1, we see that the corresponding $\mathrm{ARL}_0^{*^b} \leq 12$. In the following we assume that at Phase II we have 40 porcelains color appearance data which are grouped into 8 samples of size 5, and the controlling decisions based on the above obtained control limits are shown in Table 2.

We see that, under the necessity index based possibilistic confidence level $1 - \xi = 0.916$, only no. 2 and no. 8 groups of porcelains are completely out of control, no. 4 is in control, and the remaining 5 groups of porcelains are in the state of partial in-control with different degrees.

From this example, we can see what follows.

(i) The proposed fuzzy Shewhart control chart is completely based on the bootstrap method regardless of the distribution model of the process variable. Thus, the solution for establishing the control chart for fuzzy data is completely distribution-free.

(ii) The proposed fuzzy Shewhart control chart is an extension of the chart proposed by Grzegorzewski and Hryniewicz (2000). If the distribution of underlying process variable is known to be the normal distribution, then it will be reduced to the chart of Grzegorzewski and Hryniewicz (2000). As the necessity measure of possibility theory was applied for determining the control limits, it can be viewed as a new fuzzy control chart obtained using the direct fuzzy way (cf. Cheng, 2005; Shu and Wu, 2011; Faraz and Shapiro, 2010; Gülbay and Kahraman, 2007; 2006).

(iii) The proposed Shewhart chart is more sophisticated in design than other fuzzy control charts with known distributions of the underlying process variable under a direct fuzzy way, as we have to make an additional computational effort for doing bootstrapping and analyzing its results.

Table 1. Eight groups of five-element LR-fuzzy sample data.

j/i	1	2	3
1	(6,10,3)	(9,3,6)	(8,10,12)
2	(6,8,4)	(5,10,12)	(7,6,5)
3	(5,7,8)	(6,8,7)	(7,11,5)
4	(7,4,7)	(6,9,5)	(5,7,9)
5	(7,8,9)	(6,3,4)	(4,10,3)
$\overline{\tilde{x}}_i$	(6.4,7.4,6.2)	(6.4,6.6,6.8)	(6.2,8.8,6.8)
$\delta_2^2(\cdot,\cdot)$	3.453333	12.66667	20.43992
	1.186667	7.133333	2.854262
	0.686667	0.733333	0.354262
	4.35333	2.5	0.435274
	2.886667	3.466667	13.07578

j/i	4	5	6
1	(7,6,5)	(6,3,4)	(10,3,8)
2	(8,3,2)	(8,3,8)	(9,7,6)
3	(6,5,4)	(7,9,10)	(5,3,5)
4	(6,9,10)	(6,4,3)	(4,10,8)
5	(7,9,12)	(9,8,7)	(6,8,3)
$\overline{\tilde{x}}_i$	(6.8,6.4,6.6)	(7.2,5.4,6.4)	(6.8,6.2,6.1)
$\delta_2^2(\cdot,\cdot)$	0.414829	3.26	20.93333
	6.290829	3.626667	4.066667
	2.640163	4.36	3.133333
	3.288163	4.893333	13.43333
	6.170829	2.626667	4.6

j/i	7	8
1	(11,3,2)	(6,3,4)
2	(10,8,3)	(11,5,3)
3	(4,3,5)	(12,3,2)
4	(9,8,7)	(5,10,12)
5	(8,3,2)	(4,5,6)
$\overline{\tilde{x}}_i$	(8.4,5,3.8)	(7.6,5.2,5.4)
$\delta_2^2(\cdot,\cdot)$	8.226667	3.053333
	1.126667	8.786667
	13.22667	19.45333
	3.626667	15.52
	1.326667	11.58667
$\overline{\overline{\tilde{x}}}$	(6.95, 6.375, 6)	
$\overline{\overline{s}}$	3.082319	

5. Conclusions

The f.r.v.'s in the sense of Puri–Ralescu are always distribution free. A sample from these f.r.v.'s can be viewed as fuzzy data. We consider the case of one dimension where fuzzy data may be well expressed by a LR-fuzzy set form. Using the bootstrap approach to calculate the quantile of the fuzzy mean statistic,

Table 2. Eight groups of size five-element LR- fuzzy sample data in Phase II and the control status.

j/i	1	2	3
1	(8.5,3,3.5)	(13.6,10,12)	(9.2,8,7)
2	(7.8,6,8)	(14,9,8)	(8.6,9,10)
3	(12.3,4,10)	(11,3,7)	(16,4,7)
4	(3.2,5,6)	(6,5,9)	(2,10,11)
5	(2.3,10,3)	(6.8,6.2,9)	(5,12,8)
$\bar{\tilde{x}}_i$	(6.82,5.6,6.1)	(10.28,6.64,7.8)	(8.16,8.6,8.6)
$(\bar{\tilde{x}}_i)_{0.916}$	[6.35,7.33]	[9.72,10.94]	[7.44,8.88]
Decision	Partial	Out of Control	Partial
β	0.512		0.424

j/i	4	5	6
1	(6.9,4,8)	(6,6,4)	(7.4,6.2,5)
2	(7.5,6,8)	(7,7,8)	(8.1,6.4,8)
3	(8.1,6.3,5.2)	(7,4,3)	(6.9,6,4)
4	(6.3,5.2,7)	(6.4,10,12)	(5.9,5,6)
5	(8.2,4.3,6)	(6.8,9,13)	(6.1,6,4)
$\bar{\tilde{x}}_i$	(7.4,5.16,6.84)	(6.64,7.2,8)	(6.88,5.92,5.4)
$(\bar{\tilde{x}}_i)_{0.916}$	[6.97,7.97]	[6.04,7.31]	[6.38,7.33]
Decision	In Control	Partial	Partial
β		0.5	0.512

j/i	7	8
1	(1,8,7)	(13,8.8,6.7)
2	(2,3,2)	(19,5.3,7.8)
3	(3,4,5)	(22,4.3,6.9)
4	(9,10,13)	(37,16,21)
5	(18,5,6)	(32,6,8)
$\bar{\tilde{x}}_i$	(6.6,6,6.6)	(24.6,8.08,10.08)
$(\bar{\tilde{x}}_i)_{0.916}$	[6.1,7.15]	[23.92,25.44]
Decision	Partial	Out of Control
β	0.406	

we propose a fuzzy Shewhart control chart, which has some advantages over the existing fuzzy Shewhart control charts. Therefore, our control chart is distribution-free, i.e., does not depend upon the distributional model of the underlying process variable. Moreover, the proposed control chart method can be used for designing the control chart for d-dimensional fuzzy data if the f.r.v.'s in the sense of Puri–Ralescu can be expressed in parametric forms. Some research topics for future study include the following:

(i) instead of the distance δ_2, one can use other more appropriate distances between fuzzy data in order to develop other nonparametric control charts for fuzzy data;

(ii) under the same conditions of the fuzzy data and the related distance presented in this paper, one can construct other nonparametric control charts for fuzzy data, such as, e.g., CUSUM control charts or EWMA control charts.

Acknowledgment

Dabuxilatu Wang wishes to acknowledge the support from the National Natural Science Foundation of China under the grant no. 11271096.

References

Cen, Y. (1996). Fuzzy quality and analysis on fuzzy probability, *Fuzzy Sets and Systems* **83**(2): 283–290.

Cheng, C.-B. (2005). Fuzzy process control: Construction of control charts with fuzzy numbers, *Fuzzy Sets and Systems* **154**(2): 287–303.

Couso, I., Dubois, D., Montes, S. and Sanchez, L. (2007). On various definitions of the variance of a fuzzy random variable, *Proceedings of the 5th International Symposium on Imprecise Probabilities and Their Applications, Prague, Czech Republic*, `www.sipta.org/isipta07/proceedings/056.html`.

Dubois, D. and Prade, H. (1983). Ranking fuzzy numbers in the setting of possibility theory, *Information Sciences* **30**(3): 183–224.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman-Hall, New York, NY.

Faraz, A. and Shapiro, A. (2010). An application of fuzzy random variables to control charts, *Fuzzy Sets and Systems* **161**(20): 2684–2694.

Feng, Y., Hu, L. and Shu, H. (2001). The variance and covariance of fuzzy random variables and their applications, *Fuzzy Sets and Systems* **120**(3): 487–497.

Féron, R. (1976). Ensembles aléatoires flous, *Comptes Rendus de l'Academie des Sciences Serie A* **282**: 903–906.

Gil, M., López-Diaz, M. and Ralescu, D.A. (2006). Overview on the development of fuzzy random variables, *Fuzzy Sets and Systems* **157**(19): 2546–2557.

Grzegorzewski, P. and Hryniewicz, O. (2000). Soft methods in statistical quality control, *Control and Cybernetics* **29**(1): 119–140.

Gülbay, M. and Kahraman, C. (2006). Development of fuzzy process control charts and fuzzy unnatural pattern analysis, *Computational Statistics and Data Analysis* **51**(1): 433–451.

Gülbay, M. and Kahraman, C. (2007). An alternative approach to fuzzy control charts: Direct fuzzy approach, *Information Sciences* **177**(6): 1463–1480.

Höppner, J. (1994). *Statistische Prozeßkontrolle mit Fuzzy-Daten*, Ph.D. thesis, University of Ulm, Ulm.

Höppner, J. and Wolff, H. (1995). The design of a Shewhart control chart for fuzzy data, *Technical report*, University of Ulm, Würzburg.

Hryniewicz, O. (2006). Possibilistic decisions and fuzzy statistical tests, *Fuzzy Sets and Systems* **157**(19): 2665–2673.

Kanagawa, A., Tamaki, F. and Ohta, H. (1993). Control charts for process average and variability based on linguistic data, *International Journal of Production Research* **2**(4): 913–922.

Körner, R. (1997). On the variance of fuzzy random variables, *Fuzzy Sets and Systems* **92**(1): 83–93.

Körner, R. (2000). An asymptotic α-test for the expectation of random fuzzy variables, *Journal of Statistical Planning and Inference* **83**(2): 331–346.

Kruse, R. and Meyer, K. (1987). *Statistics with Vague Data*, D. Riedel, Dordrecht.

Kwakernaak, H. (1978). Fuzzy random variables, Part I: Definitions and theorems, *Information Sciences* **15**(1): 1–15.

Kwakernaak, H. (1979). Fuzzy random variables, Part II: Algorithms and examples for the discrete case, *Information Sciences* **17**(3): 253–278.

Liu, R. and Tang, J. (1996). Control charts for dependent and independent measurements based on bootstrap methods, *Journal of American Statistical Association* **91**(436): 1694–1700.

Montenegro, M., Colubi, A., Casals, M. and Gil, M. (2004). Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable, *Metrika* **59**(1): 31–49.

Näther, W. (2000). On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data, *Metrika* **51**(3): 201–221.

Näther, W. (2006). Regression with fuzzy random data, *Computational Statistics and Data Analysis* **51**(1): 235–252.

Nelson, L. (1985). Interpreting Shewhart $\bar{X}$ control chart, *Journal of Quality Technology* **17**(2): 114–116.

Puri, M. and Ralescu, D. (1986). Fuzzy random variables, *Journal of Mathematical Analysis and Applications* **114**(2): 409–422.

Raz, T. and Wang, J. (1990). Probabilistic and membership approaches in the construction of control charts for linguistic data, *Production Planning & Control* **1**(3): 147–157.

Senturk, S. and Erginel, N. (2009). Development of fuzzy $\widetilde{\overline{\overline{x}}} - \widetilde{R}$ and $\widetilde{\overline{\overline{x}}} - \widetilde{S}$ control charts using α-cuts, *Information Sciences* **179**(10): 1542–1551.

Shu, M.-H. and Wu, H.-C. (2011). Fuzzy $\overline{X}$ and r control charts: Fuzzy dominance approach, *Computers & Industrial Engineering* **61**(3): 676–685.

Taleb, H. (2009). Control charts applications for multivariate attribute processes, *Computers & Industrial Engineering* **56**(1): 399–410.

Taleb, H. and Limam, M. (2002). On fuzzy and probabilistic control charts, *International Journal of Production Research* **40**(12): 2849–2863.

Wang, J. and Raz, T. (1990). On the construction of control charts using linguistic variables, *International Journal of Production Research* **28**: 477–487.

Wetherill, B. and Brown, D. (1991). *Statistical Process Control*, Chapman and Hall, London.

Woodall, W.H. Tsui, K.-L. and Tucker, G.R. (1997). A review of statistical and fuzzy quality control charts based on categorical data, *in* H.-H. Lenz and P.-Th. Wilrich (Eds.), *Frontiers in Statistical Quality Control 5*, Physica-Verlag, Heidelberg, pp. 83–89.

Zadeh, L. (1965). Fuzzy sets, *Information and Control* **8**(3): 338–353.

Zadeh, L. (1975). The concept of a linguistic variable and its application to approximate reasoning, Parts 1 and 2, *Information Sciences* **8**(3): 199–249, **8**(4): 301–357.

Dabuxilatu Wang is now working as a professor at Guangzhou University, PR China. He was conferred a Bachelor's degree in mathematics by Inner Mongolia University in 1982, and a Ph.D. degree by Chiba University (Japan) in 2003. He worked as a lecturer and an associate professor in Inner Mongolia University for Nationalities from 1982 to 2003, and visited the University of Ulm, Germany, from 1998 to 1999. His research areas include statistical process control, actuarial and insurance management, and statistical analysis of fuzzy data. He has authored more than 30 publications in English and Chinese.

Olgierd Hryniewicz is a professor at the Systems Research Institute of the Polish Academy of Sciences in Warsaw. He received his Ph.D. and D.Sc. degrees in 1976 and 1985, respectively. His main area of research is statistical data analysis as well as data mining for uncertain and imprecise data. The main area of applications of his results covers quality control and reliability. He has published more than 200 papers in journals, books and conference proceedings.

Received: 24 October 2013
Revised: 2 August 2014