Fusion Method for True Value Estimation Based on Information Poor Theory

Xintao Xia
College of Mechatronical Engineering, Henan University of Science and Technology, Luoyang, China
Email: xiaxt1957@163.com

Leilei Gao and Jianfeng Chen
College of Mechatronical Engineering, Henan University of Science and Technology, Luoyang, China
Email: gll1986313@126.com, chen_jian_feng_5@163.com

***Abstract*—Point estimation is an important issue in data processing. Classical statistics is taken into account to assess mainly the true value of a population under the condition of large sample sizes and known probability distributions. If sample sizes are small and probability distributions are unknown, many statistical methods may become ineffective. For this end, a fusion method based on the information poor theory is proposed. Synthetically considering the characteristics of the various methods and effectively depicting the true value from different aspects, the fusion method is able to obtain the final estimated true value used as the most appropriate representative of the true value of a population. The results of the Monte Carlo simulation and the experimental investigation indicate that the fusion method allows the number of the data to be little and the probability distribution to be unknown.**

***Index Terms*—information fusion, point estimation, data series, small sample, information poor system**

I. Introduction

Point estimation is an important issue in data processing. Classical statistics is taken into account to assess mainly the true value of a population under the condition of large sample sizes and known probability distributions [1-3]. Commonly, the mathematical expectation or the arithmetic mean in statistics can be employed to evaluate the true value [4-6].

Nevertheless, in fields of science and technology, many problems are characterized by small sample sizes and unknown probability distributions. For example, in development and manufacture of special type bearings such as bearings for space applications, much less bearings can be made a trial due to many species, small quantity, or high cost, every time only ten-odd or even several bearings are required, thus only smaller characteristic data can be obtained. Accordingly, under the circumstances, point estimation becomes a commonly admitted difficult task. With the development of the information theory, it is possible to work it out with the help of the concept of information fusion [7-14].

In the information theory, this problem belongs to the category of the information poor system [7]. Poor information means incomplete and insufficient information, such as, in system analysis, a known probability distribution only with a small sample, an unknown probability distribution only with several data, and trends without any prior knowledge.

At present, parameter estimation for an information poor system is one of the hot topics in the field of information science and system science [1,7]. For example, Ah-Pine Julien [9] presented the fusion method for information retrieval using different aggregation operators, Zhi [10] made a damage effect assessment for battlefield target based on the multiple neural network fusion algorithm, and Wu [12] studied oil quality evaluation system and its implementation based on multi-information fusion.

The available methods for information fusion are based on the given rules which usually deal with the distribution characteristics of a data series. Although these methods can be employed to estimate the true value, different methods have different results owing to different rules. This does not mean that one method is better than another. It just shows that this method may be better for one data series, but the other method may be better for another data series. This means that the result achieved using single method is not the most reasonable "representative" of the true value based on the fact that it reflects single aspect of the attribute of a data series with poor information. For this reason, this paper develops five methods, three concepts, and one rule. The methods include the rolling mean method, the membership function method, the maximum membership grade method, the moving bootstrap method, and the arithmetic mean method. The concepts comprise the solution set on the estimated true value, the fusion series, and the final estimated true value. And the rule is the range rule. Based on this, a fusion method for the multiple estimated true values is presented to put in practice the point estimation under the condition of poor information. And the Monte Carlo simulation and the experimental investigation are done to ensure the reliability of the fusion method proposed in this paper.

Manuscript received March 1, 2011; revised June 1, 2011; accepted July 1, 2011.
Corresponding author: Xintao Xia.

doi:10.4304/jsw.7.5.1014-1021

II. Concept of Fusion for True Value

In the true value estimation, so far as the same data series with poor information is concerned, various methods should be used in order to obtain more varied information that can be utilized to describe the characteristics of the population of this data series from more aspects. As different methods have different rules [7], the characteristic information obtained is different. The characteristic information obtained has the relationship with the true value and can therefore form a set of information of the estimated true values. It is obvious that from different aspects, the set is an embodiment of the characteristics of the population. According to the information fusion theory, the true value can be evaluated by means of the set. This is the fusion method for the multiple estimated true values.

Suppose a raw data series, viz., a small sample of size N, outputted by a population, can be given by

$$X = (x(1), x(2), \cdots, x(t), \cdots, x(N)); t = 1,2,\cdots, N; N > 2 \quad (1)$$

where t is the sequence number, $x(t)$ is the tth data in X, and N is the number of the data.

It must be pointed out that according to the information poor theory, N can be a small real number and the probability distribution of X can be unknown.

Definition 1: Solution set is a set on the true value estimated using the L methods. It is noted as

$$X_0 = (X_{01}, X_{02}, \cdots, X_{0l}, \cdots, X_{0L}); l = 1,2,\cdots, L \quad (2)$$

where X_0 is called the solution set and X_{0l} is called the lth estimated true value obtained using the lth method in mathematics.

The solution set is not the final solution to the problem. It is only a set on the varied characteristic information of the population.

Definition 2: Fusion for the multiple estimated true value means that a new value related with the set X_0 can be acquired via a special method of mathematical treatment. This new value is the final solution $X_{0\mathrm{True}}$ in accord with a rule Θ, which is denoted by

$$X_{0\mathrm{True}} \,|\, \Theta \,|\, \mathrm{Fusion}\; X_0 \subseteq A_{\mathrm{True}} \quad (3)$$

where $X_{0\mathrm{True}}$ is the final solution that is a estimation for the true value of the population, viz., the final estimated true value; A_{True} is the true value set on the varied attributes of the population; symbol $|\Theta$ is according to the rule Θ; symbol $|\mathrm{Fusion}\; X_0$ means that X_0 is fused using the L methods in mathematics; and symbol $X_0 \subseteq A_{\mathrm{True}}$ means that X_0 is included in A_{True}.

Fusion for the true value, in fact, is that through the mathematical treatment of the solution set X_0, a finally estimated true value is obtained, which is the final estimation for the true value of the attribute of the population.

In Definition 1, the L methods in mathematics are usually based on the information poor theory. The information poor theory is a theory on analysis for a data series with poor information and consists of many mathematical theories which are different from classical statistics. For example, the fuzzy set theory [5], the grey system theory, the bootstrap methodology [7], the chaotic theory [15], and the Bayesian statistics belong to the category of the information poor theory. In point estimation, many concepts and methods in the information poor theory can be employed to evaluate the true value. This paper introduces the five methods, as shown in Table 1, for fusion of the true value.

TABLE I.
Main Method Adopted in Information Poor Theory

No. l	Method	Remark
1	Rolling mean method	Variation of weighted average method
2	Membership function method	Method based on fuzzy set theory
3	Maximum membership grade method	Method based on fuzzy set theory
4	Moving bootstrap method	Variation of bootstrap methodology
5	Arithmetic mean method	Method based on statistical theory

III. Method for Estimated True Value

There are many methods for point estimation in statistics, each of which can be used to obtain the estimated true value, but not all the methods are suitable to analyze an information poor system because of their specified requirements for prior knowledge. For example, the maximum-likelihood method requires a known probability distribution, the Bayesian method exacts a prior distribution, and the histogram method is based on a number of data and so on. At this point, the five methods shown in Table 1 may be suitable to settle the problem on both a small sample size and an unknown probability distribution.

The reason that only five methods are considered is that the more methods used, the more the prior knowledge required, but the fewer methods used, the fewer the characteristic information obtained. Commonly the number of the methods used can be 4-6 according to the information poor theory.

The reason for using the five methods in Table 1 is connected with their characteristics and is explained in detail below.

A. Rolling Mean Method

The 1st method is called the rolling mean method. It is a cross between the weighted average method and the bootstrap resampling.

Rearrange $x(t)$ taken from (1) from small to large to form a new order as

$$x_i \le x_{i+1}; i = 1,2,\cdots, N-1 \quad (4)$$

An estimated true value of the population is defined as

$$X_{01} = \sum_{j=1}^{N} \omega_j x_j = \frac{1}{N}\sum_{j=1}^{N} \xi_j \quad (5)$$

with

$$\xi_j = \frac{1}{N-j+1}\sum_{i=1}^{N-j+1}\sum_{k=i}^{i+j-1}\frac{x_k}{j}; j = 1,2,\cdots,N \quad (6)$$

where X_{01} is the 1st estimated true value, ω_j is the weight of the data x_j, and ξ_j is the jth factor of the rolling mean.

The essential thinking of the rolling mean method roots in the bootstrap resampling. With the help of the resampling in sequence, the number of the data sampled each time changes and increases in the range from 1 to N, together with calculating the factor of the rolling mean. Thus the estimated true value is approaching the true value of the population step by step.

The rolling mean method belongs to the weighted average method which has the fixed weight sequence. And the closer the distance between the data x_j and the estimated true value X_{01} is, the larger the weight of the data x_j is.

The reason to use the fixed weight sequence is due to the lack of prior knowledge of the weight in information poor system analysis.

B. Membership Function Method

The 2nd method is called the membership function method. It is based on the membership function in the fuzzy set theory.

By means of (1) and (4), define the difference sequence as

$$d = (d_1, d_2, \cdots, d_i \cdots, d_{N-1}) \quad (7)$$

with

$$d_i = x_{i+1} - x_i \ge 0; i = 1,2,\cdots,N-1 \quad (8)$$

where d_i is called the difference between x_{i+1} and x_i, the difference value for short.

From the point of view of the possibility theory, X in (1) can be considered as an event. The smaller the difference value d_i is, the thicker the distribution of the data x_i is, indicating that the possibility of the event in the interval $[x_i, x_{i+1}]$ is likely to be large. The larger the difference value d_i is, the thinner the distribution of the data x_i is, indicating that the possibility of the event in the interval $[x_i, x_{i+1}]$ is likely to be small. That is to say, the difference value d_i is closely related to the distribution density. Accordingly, define the linear membership function f_i as

$$f_i = 1 - \frac{d_i - d_{\min}}{d_{\max}}; i = 1,2,\cdots,N-1 \quad (9)$$

with

$$d_{\min} = \min_{i=1}^{N-1} d_i \quad (10)$$

and

$$d_{\max} = \max_{i=1}^{N-1} d_i \quad (11)$$

where $d_{\min}$ and $d_{\max}$ are called the minimum difference value and the maximum difference value, respectively.

Set the immediate neighbor mean series is

$$Z = (z_1, z_2, \cdots, z_i, \cdots z_{N-1}) \quad (12)$$

with

$$z_i = \frac{1}{2}(x_{i+1} + x_i); i = 1,2,\cdots,N-1 \quad (13)$$

where z_i is the ith immediate neighbor mean.

An estimated true value of the population is defined as

$$X_{02} = \frac{1}{\sum_{i=1}^{N-1} f_i}\sum_{i=1}^{N-1} f_i z_i \quad (14)$$

where X_{02} is the 2nd estimated true value.

It can be seen from (9) - (14) that the membership function method also belongs to the weighted average method which has the varying weight sequence formed by the membership function f_i. And the smaller the difference value d_i is, the larger the weight of the ith immediate neighbor mean z_i is.

The reason to use the varying weight sequence is that the weight sequence can be automatically recognized according to the discrete characteristics of the raw data series, without any prior knowledge of the weight in information poor system analysis.

C. Maximum Membership Grade Method

The 3rd method is called the maximum membership grade method. It is based on the principle of the maximum membership grade in the fuzzy set theory.

Via (9), let the maximum membership grade $f_{\max}$ be

$$f_{\max} = \max_{j=1}^{N-1} f_j = 1 \quad (15)$$

Define the mean of the two data x_{v+1} and x_v corresponding to the maximum membership grade $f_{\max}$ as an estimated true value of the population, as follows:

$$X_{03} = \frac{1}{2}(x_{v+1} + x_v); v \in (1,2,\cdots,N-1) \quad (16)$$

If there are T repeated maximum membership grades, let the τth mean be

$$x_{0\tau} = \frac{1}{2}(x_{v+1} + x_v)_\tau; \tau = 1,2,\cdots,T-1 \quad (17)$$

An estimated true value of the population is defined as

$$X_{03} = \frac{1}{T-1}\sum_{\tau=1}^{T-1} x_{0\tau} \quad (18)$$

where τ stands for the position of the τth data corresponding to the maximum membership grade.

The maximum membership grade method is a special mean method, in which the mean of the several data, in general the two data, only corresponding to the minimum difference value $d_{\min}$ are used to calculate the estimated true value.

The reason for only using the several data is that they are closely concerned with the true value according to the principle of the maximum membership grade in the fuzzy set theory. The smaller the minimum difference value $d_{\min}$, the better the several data conform to the true value, indicating that the relative error between the estimated true value and the true value is likely to be small. The larger the minimum difference value $d_{\min}$, the worse the several data conform to the true value, indicating that the relative error between the estimated true value and the true value is likely to be large.

D. Moving Bootstrap Method

The 4th method is called the moving bootstrap method. It is an improved bootstrap method, with the feature of dynamic evaluation.

For the convenience, a data series is represented as a vector, thus (1) is written by

$$\mathbf{X} = \{x(t); t = 1,2,\cdots,N\} \tag{19}$$

In (19), the sequence number t can be regarded as the time t. This has no effect on the result estimated for the true value.

In a measuring process, the problem studied with the moving bootstrap method is an estimation of the true value at the time t by means of the m data, which are close and before the time t (including the time t). And the parameter m is a very small integer. The smaller the value of the parameter m is, the fresher the information taken from the population is. In addition, the reason that only m data are taken into account at the time t is that the estimated true value will change with time in the measuring process, in which the larger the value m is, the larger the errors estimated for the true value is. According to the bootstrap methodology, the value of the parameter m is less than or equal to N.

At the time t, the m data taken from the series $\boldsymbol{X}$ can form a sub-series X_m as follows:

$$\mathbf{X}_m = \{x_m(u)\}; u = t-m+1, t-m+2,\cdots,t; t \ge m \tag{20}$$

where u is the time.

According to the bootstrap methodology, B simulation samples of size m, namely the bootstrap resampling samples, can be obtained by an equiprobable sampling with replacement from (20), as follows:

$$\mathbf{Y}_{\text{Bootstrap}} = (\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_b, \cdots, \mathbf{Y}_B) \tag{21}$$

with

$$\mathbf{Y}_b = \{y_b(u)\}; b = 1,2,\cdots,B \tag{22}$$

where $\boldsymbol{Y}_b$ is the bth bootstrap sample, $y_b(u)$ is the uth bootstrap resampling sample within $\boldsymbol{Y}_b$, and B is the number of the bootstrap resampling samples. The mean of $\boldsymbol{Y}_b$ is given by

$$y_{mb} = \frac{1}{m}\sum_{u=t-m+1}^{t} y_b(u); t = m, m+1,\cdots,N \tag{23}$$

Thus at the time w=t+1, a sample of size B can be obtained, as follows:

$$\mathbf{X}_{mw} = \{y_{mb}(w)\}; b = 1,2,\cdots,B; w = t+1 \tag{24}$$

Using (24), a probability function can be obtained as

$$F_{wB} = F_{wB}(x_m) \tag{25}$$

According to the histogram principle in statistics, the true value at the time w can be estimated by a weighted mean, as follows:

$$X_{04}(w) = \sum_{q=1}^{Q} F_{wBq} x_{mq} \tag{26}$$

where X_{04} (w) is the estimated true value at the time w; Q is the number of groups; q is the qth group, q=1,2,…, Q; x_{mq} is the median of the qth group; and F_{wBq} is the value of the bootstrap probability at the point x_{mq}.

An estimated true value of the population is define as

$$X_{04} = \frac{1}{N-m+1}\sum_{k=m}^{N} X_{04}(k) \tag{27}$$

The reason to use the moving bootstrap method is that unknown probability distributions are simulated and then mathematical expectation is obtained only using very little data.

E. Arithmetic Mean Method

The 5th method is the arithmetic mean method. It is one of most commonly used methods for point estimation. In this paper, the reason to use this method is that the data are generally observed in an equal weight manner and should be equally treated if the lack of prior knowledge on the probability distribution of the population. Therefore, this method is characterized by the equal weight means.

An estimated true value of the population is defined as

$$X_{05} = \frac{1}{N}\sum_{t=1}^{N} x(t) \tag{28}$$

It is obvious that the property of the five methods above is different. Accordingly, they can be applied to mine five different kinds of feature information of the population by processing the raw data series X, laying the foundation for the fusion method for the multiple estimated true values [7,9,10,12].

IV. Fusion Method for Estimated True Value

In order to evaluate the final result, the solution set X_0 obtained using the five methods can be regarded as a data series. It is defined as the 0th fusion series, as follows:

$$X_{0\text{Fusion}0} = (X_{01\text{F}0}, X_{02\text{F}0}, \cdots, X_{0l\text{F}0}, \cdots, X_{0L\text{F}0}) = (X_{01}, X_{02}, \cdots, X_{0l}, \cdots, X_{0L}) \quad (29)$$

with

$$X_{0l\text{F}0} = X_{0l} \quad (30)$$

where $X_{0l\text{F}0}$ is the lth fused true value calculated using the lth method to process the raw data series X.

The 0th fusion series $X_{0\text{Fusion}0}$ is fused with the help of the five methods and then a new data series, which is called the 1st fusion series $X_{0\text{Fusion}1}$, is obtained as

$$X_{0\text{Fusion}1} = (X_{01\text{F}1}, X_{02\text{F}1}, \cdots, X_{0l\text{F}1}, \cdots, X_{0L\text{F}1}) \quad (31)$$

where $X_{0l\text{F}1}$ is the lth fused true value calculated using the lth method to process the 0th fusion series $X_{0\text{Fusion}0}$.

The 1st fusion series $X_{0\text{Fusion}1}$ is fused with the help of the five methods and then a new data series, which is called the 2nd fusion series $X_{0\text{Fusion}2}$, is obtained, as follows:

$$X_{0\text{Fusion}2} = (X_{01\text{F}2}, X_{02\text{F}2}, \cdots, X_{0l\text{F}2}, \cdots, X_{0L\text{F}2}) \quad (32)$$

where $X_{0l\text{F}2}$ is the 2nd fused true value calculated using the lth method to process the 1st fusion series $X_{0\text{Fusion}1}$.

In the same way, the jth fusion series $X_{0\text{Fusion}j}$ is obtained as

$$X_{0\text{Fusion}j} = (X_{01\text{F}j}, \cdots, X_{0l\text{F}j}, \cdots, X_{0L\text{F}j});\ j = 1,2,\cdots \quad (33)$$

where $X_{0l\text{F}j}$ is the jth fused true value calculated using the lth method to process the j-1th fusion series $X_{0\text{Fusion}j\text{-}1}$.

If the jth fusion series $X_{0\text{Fusion}j}$ is in accord with the given rule Θ, the final estimated true value $X_{0\text{True}}$ is acquired.

Rule 1: Define the range as

$$\delta_j = \max_{l=1}^{L} X_{0l\text{F}j} - \min_{l=1}^{L} X_{0l\text{F}j} \quad (34)$$

If $\delta_j \le \varepsilon$ where ε is a arbitrarily small positive number, the final estimated true value $X_{0\text{True}}$ is obtained by

$$X_{0\text{True}} = \frac{1}{L}\sum_{l=1}^{L} X_{0l\text{F}j} \quad (35)$$

This is the rule Θ. It is called the range rule and, in fact, is a convergence criterion, in which the estimated true values converge to a point, i.e., the true value.

V. Experimental Investigation

The experiment is done to ensure the reliability of the fusion method proposed in this paper. The main points of the plan and thought of the experiment are as follows:

1. Acquiring conventional true value

In the experiment, lots of data must be collected in order to acquire the conventional true value. These data form a data series, viz., a large sample of size N_{Test}, which is called the testing series, denoted by X_{Test}. Using the five methods in Table 1 to process the testing series, five estimated true values, denoted by $X_{\text{m}01}$, $X_{\text{m}02}$, $X_{\text{m}03}$, $X_{\text{m}04}$, and $X_{\text{m}05}$, respectively, based on the large sample size, can be obtained.

According to the law of large numbers and the central limit theorem in classical statistics, each of the five estimated true values should equal or be approaching a fixed constant called the mathematical expectation E_{m} because of the large sample sizes, i.e., the five estimated true values should be equal to each other in theory. But the difference, in fact, always appears among them due to much interference in the process of measurement and the additional uncertainty introduced by finite sampling. According to the error theory and the statistical theory, if the difference is very small, the mean of the five estimated true values based on the large sample size can be considered as approaching the mathematical expectation E_{m} and can therefore be regarded as a conventional true value X_{True}, viz., there is

$$X_{\text{m}} = \frac{1}{5}\sum_{l=1}^{5} X_{\text{m}0l} = X_{\text{True}} \subseteq A_{\text{True}} \quad (36)$$

where X_{m} stands for the mean of the five estimated true values based on the large sample size, the mean for short, and X_{True} stands for the conventional true value.

2. Making small sample

Choose the former N data in the testing series X_{Test} as the raw data series X, where N is the number of data in X and is a very small integer for the condition of the small sample size. As a rule, N is selected in the range from 3 to 10 according to the information poor theory.

3. Calculating final estimated true value

By means of (1)-(35), the final estimated true value $X_{0\text{True}}$ is obtained. It is the estimation for the true value of the testing series X_{Test}.

4. Testing estimated result

The relative error E_{R} of estimation can be used for testing the estimated result, as follows:

$$E_{\text{R}} = \left|\frac{X_{0\text{True}} - X_{\text{True}}}{X_{\text{True}}}\right| \times 100\% \quad (37)$$

A. Testing Series

The experimental investigation deals with information of the quality parameters of the tapered roller bearing marked with 30204. The manufacturing quality parameters considered are the raceway roundness and the rib surface roughness of the inner ring. Thus the two testing series are obtained in the experimental investigation, each of which has thirty data, as shown in Fig. 1.

B. Conventional True Value Based on Large Sample Size

As described above, the mean X_{m} can be employed for expressing the conventional true value X_{True} under the condition of the large sample sizes. Based on this and considering the thirty data as a large sample of size N_{Test}=30, the five estimated true values can be calculated using the five methods, respectively. The results are shown in Figs 2 and 3.

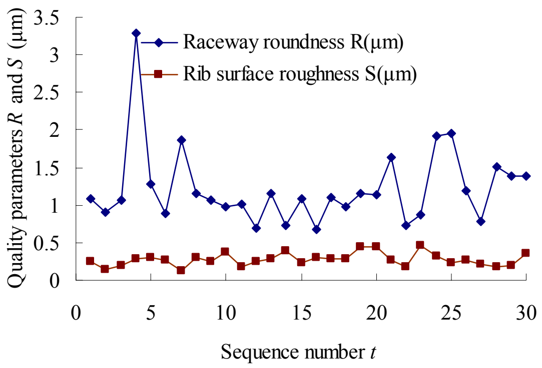

Figure 1. Two testing series obtained in experimental investigation.

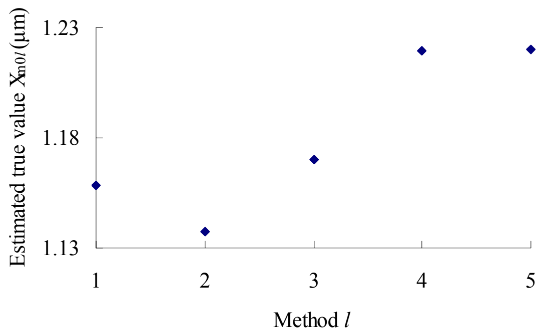

Figure 2. Estimated true value of information of raceway roundness based on large sample of size N_{Test}=30

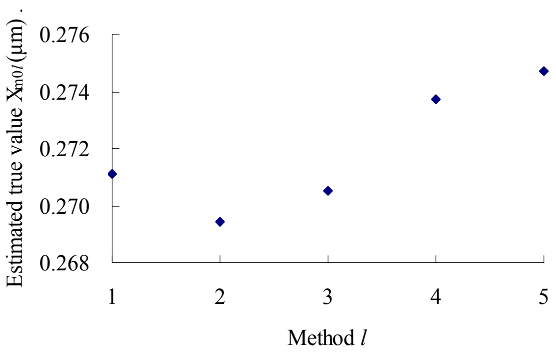

Figure 3. Estimated true value of information of rib surface roughness based on large sample of size N_{Test}=30

It can be seen from Fig. 2 that for the raceway roundness based on the large sample size, the five estimated true values take values in the range from X_{m02}=1.13727 μm to X_{m05}=1.22 μm, only with a very small difference between X_{m02}=1.13727 μm and X_{m05}=1.22 μm (the relative error is only 0.7%). The mean of the five estimated true values in Fig. 2 is equal to 1.181 μm. It can be regarded as the conventional true value X_{True} of the raceway roundness of the inner ring.

It can be seen from Fig. 3 that for the rib surface roughness based on the large sample size, the five estimated true values take values in the range from X_{m02}=0.26944 μm to X_{m05}=0.27473 μm, only with a very small difference between X_{m02}=0.26944 μm and X_{m05}=0.27473 μm (the relative error is only 1.9%). The mean of the five estimated true values in Fig. 3 is equal to 0.2719 μm. It can be regarded as the conventional true value X_{True} of the rib surface roughness of the inner ring.

C. Evaluation of Information of Raceway Roundness Based on Small Sample Size

From Fig. 1, choose the former 5 data (viz., a small sample of size N=5) in the testing series X_{Test} about the raceway roundness R as the raw data series X=(1.08, 0.9, 1.06, 3.28, 1.28). Using the five methods in Table 1 to process the raw data series X, the results are shown in Table 2.

TABLE II.
ESTIMATED TRUE VALUE OF INFORMATION OF RACEWAY ROUNDNESS BASED ON SMALL SAMPLE OF SIZE N=5

No. l	Method	Estimated true value X_{0l}[μm]	Relative error E_R(%)
1	Rolling mean method	1.428	20.91
2	Membership function method	1.080	8.55
3	Maximum membership grade method	1.070	9.40
4	Moving bootstrap method	1.578	33.62
5	Arithmetic mean method	1.520	28.70

It is easy to see from Table 2 that the five results are different each other in numerical value. The estimated true value takes values in the range from 1.070 μm to 1.578 μm. And the minimum and the maximum of the relative errors are 8.55% and 33.62%, respectively.

According to Define 1, the solution set X_0 can be formed by the estimated true value X_{0l} in Table 2, viz., X_0= (1.428, 1.080, 1.070, 1.578, 1.520).

It is to be noted that the solution set X_0 obtained using the five methods is not the final solution to the problem. It is only a set on the varied characteristic information of the population. In order to obtain the final result, let the 0th fusion series be $X_{0\text{Fusion}1}$=X_0. Then using the rolling mean method to process the 0th fusion series $X_{0\text{Fusion}0}$, the 1st fused true value 1.339 μm can be obtained. Using the membership function method to process the 0th fusion series $X_{0\text{Fusion}0}$, the 2nd fused true value 1.346 μm can be obtained. Using the maximum membership grade method to process the 0th fusion series $X_{0\text{Fusion}0}$, the 3rd fused true value 1.075 μm can be obtained. Using the moving bootstrap method (let B=50000, Q=8, and m=4) to process the 0th fusion series $X_{0\text{Fusion}0}$, the 4th fused true value 1.271 μm can be obtained. Using the arithmetic mean method to process the 0th fusion series $X_{0\text{Fusion}0}$, the 5th fused true value 1.335 μm can be obtained.

These five fused true values can form the 1st fusion series $X_{0\text{Fusion}1}$=(1.339, 1.346, 1.075, 1.271, 1.335).

In the same way, many fusion series are obtained successively. Let ε=0.001, according to Rule 1, the final estimated true value $X_{0\text{True}}$ is 1.3. And the relative error is 10.08%.

D. Evaluation of Information of Rib Surface Roughness Based on Small Sample Size

From Fig. 1, choose the former 5 data (viz., a small sample of size N=5) in the testing series X_{Test} about the raceway roundness R as the raw data series X=(0.247, 0.148, 0.197, 0.276, 0.306).

The estimated true values of the rib surface roughness based on the small sample size are shown in Table 3. It is easy to see from Table 3 that the five results are different each other in numerical value. The estimated true value takes values in the range from 0.2180 μm to 0.2615 μm. Let ε=0.001. According to Rule 1, the final estimated true value $X_{0\text{True}}$ is 0.2391 and the relative error is 12.06%.

TABLE III.
ESTIMATED TRUE VALUE OF INFORMATION OF RIB SURFACE ROUGHNESS BASED ON SMALL SAMPLE OF SIZE N=5

No. l	Method	Estimated true value X_{0l}[μm]	Relative error E_R(%)
1	Rolling mean method	0.2362	13.13
2	Membership function method	0.2465	9.34
3	Maximum membership grade method	0.2615	3.82
4	Moving bootstrap method	0.2180	19.82
5	Arithmetic mean method	0.2348	13.64

E. Monte Carlo Simulation

A stochastic process of the uniform distribution (with the [0,1] interval) is simulated using the Monte Carlo method, then the testing series X_{Test} is taken from the stochastic process (as shown in Fig.4, the number of the data in the testing series X_{Test} is N_{Test}=1024). Clearly, the testing series X_{Test} conforms to the uniform distribution. According to statistics, the mathematical expectation E_m of the uniform distribution is (1−0)/2=0.5 and hence the conventional true value of the testing series X_{Test} is X_{True}=0.5. Choose the former 5 data (a small sample of size N=5) in the testing series X_{Test} about the uniform distribution U as the raw data series X=(0.77374, 0.40697, 0.72449, 0.55777, 0.38796). The estimated true values of the uniform distribution based on the small sample size are shown in Table 4.

It is easy to see from Table 4 that the five results are different each other in numerical value. The estimated true value takes values in the range from 0.39746 to 0.58083. And the minimum and the maximum of the relative errors are 10.62% and 20.51%, respectively.

Let ε=0.0005. According to Rule 1, the final estimated true value $X_{0\text{True}}$ is 0.55066 and the relative error is 10.13%.

Under the condition of the small sample size, the relative errors between the final estimated true value and the conventional true value take values in the range from 10.08% to 12.06%. Using the fusion method, the maximum of the relative errors is 12.06%, which is smaller than that can be produced by the other five methods.

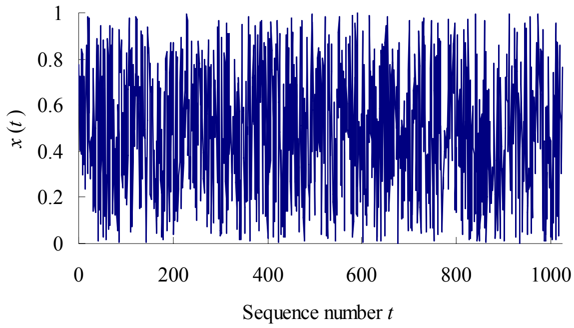

Figure 4. Testing series of uniform distribution

TABLE IV.
ESTIMATED TRUE VALUE OF UNIFORM DISTRIBUTION BASED ON SMALL SAMPLE OF SIZE N=5

No. l	Method	Estimated true value X_{0l}	Relative error E_R(%)
1	Rolling mean method	0.56836	13.67
2	Membership function method	0.58083	16.17
3	Maximum membership grade method	0.39746	20.51
4	Moving bootstrap method	0.55312	10.62
5	Arithmetic mean method	0.57018	14.04

VI. DISCUSSIONS

As described above, the fusion method proposed in this paper contains three steps. First, for a raw data series outputted by a population, the solution set structured with different estimated truth values is obtained via different methods with different properties. Second, a chain of the fusion series composed of the fused true values is formed using different methods to process repeatedly the fusion series. Finally, the final estimated truth value is accepted according to the range rule, which is the appropriate estimation for the true value of the population. In this way, the relative error of estimation is very small. For example, in the experimental investigation on the quality parameters of the tapered roller bearing marked with 30204, the maximum of the relative errors is only 12.06%. Therefore, the result estimated using the fusion method presented in this paper is suitable for the reasonable "representative" under the condition of poor information.

It is found through the experimental investigation that the arithmetic mean method in statistics is not the best. Sometimes, it is able to bring a very large estimated error. For example, the relative error is up to 28.70% when the raceway roundness is evaluated, indicating the statistical method is likely to be inappropriate under the condition of small sample sizes.

It is reported that the data series about the roundness and the roughness possess different probability distributions. Therefore, it will be seen from the experimental investigation that the fusion method for the estimated true value presented in this paper is always the appropriate method in evaluating the quality parameters which conform to different probability distributions.

According to Figs. 2 and 3, under the condition of large sample sizes (with the thirty data in the testing series), the five results obtained using the five methods proposed in Table 1 are close to each other in numerical value. But according to Tables 2 to 4, under the condition of small samples (only with the five data in the raw data series *X*), the difference among the five results is very large. This means that the five methods in Table 1 are effective under the condition of large samples (for example as many as thirty data), but they are likely to be ineffective under the condition of small samples (for example as few as five data). Nevertheless, under the condition of the small sample size, if the five results are processed using the fusion method developed in this paper, the good results can be attained and the maximum of the relative errors between the final estimated true value and the conventional true value is only 12.06%. It is follows that the fusion method for the estimated true value can be one of complements for the available statistical methods and data fusion methods in use.

In addition, when the true value is estimated under the condition of small sample sizes and unknown probability distributions, if the fusion method is utilized, the maximum of the relative errors is commonly small, but if single method is utilized, the maximum of the relative errors is commonly large.

As a result, the method for fusion of methods is good at solving the problem on both small sample sizes and unknown probability distributions, with the reliable estimated result.

VII. CONCLUSIONS

The fusion method for the multiple estimated true values allows the number of the data to be very little and the probability distribution to be unknown.

The fusion method is able to obtain the final estimated true value that can be used as the most appropriate representative of the true value of the information poor system.

The fusion method for the multiple estimated true values can be one of complements for the available statistical methods and data fusion methods in use.

ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China (Grant No. 51075123).

REFERENCES

[1] F. Xia, H. Zhang, D. G. Peng, H. Li, and Y. K. Su, "Condenser fault diagnosis based on FNN and data fusion," *Applied Mechanics and Materials*, vols. 44–47, pp. 3762–3766, 2011.

[2] M. VanHilst, S. H. Huang, and H. Lindsa, "Process analysis of a waterfall project using repository data," *International Journal of Computers and Applications*, vol. 33, no. 1, pp. 49–56, 2011.

[3] M. V. Kulikova, "Maximum likelihood estimation via the extended covariance and combined square-root filters," *Mathematics and Computers in Simulation*, vol. 79, no. 5, pp. 1641–1657, 2009.

[4] Lingasubramanian Karthikeyan, Alam Syed M., and Bhanja Sanjukta, "Maximum error modeling for fault-tolerant computation using maximum a posteriori (MAP) hypothesis," *Microelectronics Reliability*, vol. 51, no. 2, pp. 485–501, 20116.

[5] M. N. Mokhtarian, "A new fuzzy weighted average (FWA) method based on left and right scores: An application for determining a suitable location for a gas oil station," *Computers and Mathematics with Applications*, vol. 61, no. 10, pp. 3136–3145, 2011.

[6] J. L. Liu, J. P. Li, W. X. Xu, and Y. Shi, "A weighted Lq adaptive least squares support vector machine classifiers-Robust and sparse approximation," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2253–2259, 2011.

[7] X. T. Xia, T. M. Lv, and F. N. Meng, "Gray chaos evaluation model for prediction of rolling bearing friction torque," *Journal of Testing and Evaluation*, vol. 38, no. 3, pp. 291–300, 2010.

[8] Ah-Pine Julien, "On data fusion in information retrieval using different aggregation operators," *Web Intelligence and Agent Systems*, vol. 9, no. 1, pp. 43–55, 2011.

[9] J. Cao and J. R. Zheng, "Multi-person tracking based on video information fusion," *Advanced Materials Research*, vol. 187, pp. 452–457, 2011.

[10] J. Zhi, J. Y. Liu, and C. Q. Fu, "Damage effect assessment of battlefield target based on multiple neural network fusion algorithm," *Journal of Computers*, vol. 5, no. 10, pp. 1590–1596, 2010.

[11] P. Y. Zhu, Q. Zhong, W. L. Xiong, and B. G. Xu, "A robust least squares for information estimation fusion," *International Journal of Digital Content Technology and its Applications*, vol. 5, no. 4, pp. 65–73, 2011.

[12] X. X. Wu, "Study on oil quality evaluation system and its implementation based on multi-information fusion," *Advanced Materials Research*, vols. 197–1985, pp. 1486–1493, 2011.

[13] X. F. Liu, X. Z. Ye, S. Y. Zhang, and F. Liu, "Regularized least squares estimating sensitivity for self-calibrating parallel imaging," *Journal of Computers*, vol. 6, no. 5, pp. 857–864, 2011.

[14] S. J. Li, Y. L. Zhang, M. Xie, and H. Sun, "Belief reasoning recommendation: Mashing up web information fusion and FOAF," *Journal of Computers*, vol. 5, no. 12, pp. 1885–1892, 2010.

[15] X. T. Xia and T. M. Lv, "Chaos prediction of rolling bearing friction torque," *Applied Mechanics and Materials*, vols. 26-28, pp. 190–193, 2010.

Xintao Xia is a professor at College of Mechatronical Engineering, Henan University of Science and Technology. He received his PhD degree in Mechanical Science Engineering from Shanghai University in 2008. His main research areas include information processing, information poor system, vibration and noise. He has published more than 170 journal papers and 11 technical monographs since 1985.