Journal of Industrial and Systems Engineering
Vol. 3, No. 1, pp 56-71
Spring 2009

A New Approach to Distribution Fitting: Decision on Beliefs

Ali Eshragh[1*], Mohammad Modarres[2]

[1] Department of Industrial Engineering, Sharif University of Technology, Tehran, IRAN; Presently at the School of Mathematics and Statistics, The University of South Australia, Mawson Lakes, SA 5095, Australia
Ali.EshraghJahromi@unisa.edu.au

[2] Department of Industrial Engineering, Sharif University of Technology, Tehran, IRAN
modarres@sharif.edu

ABSTRACT

We introduce a new approach to distribution fitting, called *Decision on Beliefs* (DOB). The objective is to identify the probability distribution function (PDF) of a random variable X with the greatest possible confidence. It is known that f_X is a member of $S = \{f_1, \cdots, f_m\}$. To reach this goal and select f_X from this set, we utilize stochastic dynamic programming and formulate this problem as a special case of Optimal Stopping Problem. The decision is made on the basis of the outcome of a limited number of experiments. A real number, namely, *belief* is assigned to each candidate by considering the outcome of observations. At each stage and after a random observation, beliefs are updated by applying Bayesian formula and then either one element of S is selected as the desired PDF or another observation is made. At each stage, a PDF from S with the greatest belief is accepted as the desired PDF provided the belief is higher than a least acceptable designated level. We assume the total number of possible observations can not exceed N and a cost is incurred for each observation. Dynamic and nonlinear programming are applied to calculate the least acceptable belief value for each stage. To reduce the search of the optimal solution, the concept of entropy is utilized.

Keywords: Distribution fitting, Dynamic programming, Markovian decision process.

1. INTRODUCTION

Identifying a suitable probability distribution function (PDF) to fit the set of data obtained from experiments is the first step in majority of statistical analysis cases. This paper introduces a new approach to distribution fitting, called *Decision on Beliefs* (DOB) (Eshragh, 2001).

There are different approaches for distribution fitting. However, *Goodness of Fit* (GOF) is the most popular one in the literature (Conover, 2001). A random sample is drawn from some population and examined some way in order to make sure the distribution fits the data reasonably. The objective of GOF is to test the null hypothesis, *i.e.* to show the unknown PDF has a known and specified distribution function. In every GOF test, all distribution candidates must be checked one by one suitability, regardless of their effects on each other. Consequently, more than one distribution may

[*] Corresponding Author

be accepted as the desired distribution. Therefore, not necessarily a unique PDF is selected. In contrast, we show DOB approach does not create this difficulty and only one candidate is accepted. Furthermore, DOB is especially a useful technique if the number of experiments for identifying the distribution function is limited due to its high cost or the nature of the experiments.

It ought to be mentioned that we do not focus on the algorithmic potential of the results reported here. However, there is a great deal of debate about this topic in Saniee Monfared and Ranaeifar (2007). Our chief aim in this paper is suggesting a number of algorithmic approaches that could be the subject of continuing research.

In the following section, the problem is defined and our general approach to distribution fitting is also presented. In Section 3, we define the concept of beliefs and how to update them stochastically after receiving a new observation. In Section 4, we restrict the search region by making some pairwise comparisons. Determining the least acceptable belief is investigated in Section 5 and the algorithm is described in Section 6. To illustrate the method, a numerical example is presented in details in Section 7. In Section 8 we make conclusion and suggest some future research topics and the proof of theorems are moved to the end of paper and appear in appendices.

2. THE PROBLEM

Consider a continuous random variable X with PDF f_X and CDF F_X. Although f_X is unknown, we know it belongs to a candidate set $S=\{f_1,\cdots,f_m\}$. We assume the distribution as well as the parameters of each member of S is known. The objective is to identify f_X, *i.e.* select one element of this set as f_X with the greatest confidence within limited number of observations.

2.1. General Approach

Selecting the desired PDF from the candidate set of $S=\{f_1,\cdots,f_m\}$ is by making observations from f_X, sequentially. After each observation, a real number called *"belief"* is assigned to each member of $S=\{f_1,\cdots,f_m\}$. Actually, the belief on each PDF, say f_i, is the probability that f_i is the desired PDF, based on the information obtained from the observations up to now. After each observation, it is decided whether to select one candidate from $S=\{f_1,\cdots,f_m\}$ as f_X or continue and try another observation. Clearly, with each new observation the decision makers obtains more information and as a result the beliefs change accordingly. We also assume the total number of possible observations is limited to N and a cost is incurred for each observation.

The procedure stops and selects a PDF, say f_i, if it has the greatest belief among the candidates and its belief is not less than a predetermined value which indicates the *least acceptable belief*. The procedure also stops after N observations.

In the subsequent sections, to explain the above general approach we need to define the concept of beliefs and the way which they update. Furthermore, to determine the optimal least acceptable belief, the problem is formulated as a special case of *Optimal Stopping Problem*. To do that, we apply Markovian decision process or in fact stochastic dynamic programming, see Ross (1983).
The proposed method is easy to be implemented. However, the mathematical theorems to support this method are lengthy and to some extent complicated. Therefore, most of the proofs are presented in the appendices.

The only paper which is somehow related to this work is by Ahn and Kim (1998). They developed a method to maximize the utility obtained by observations to make the best decision regarding the action time, when observations are made sequentially.

3. THE BELIEF

Definition 1. Let the outcome of j th observation be denoted by x_j. Then, after k observations, for $k = 1,\cdots,N$, we call $O_k = (x_1,\cdots,x_k)$ as k th *observations vector*.

Definition 2. We introduce the concept of *belief* on f_i as the probability that f_i is the desired PDF, on the basis of the information obtained from the outcome of the observations up to this point (Bernardo and Smith, 2001). In other words, if the k th observations vector is O_k, then the belief on f_i is defined as follows.

$$B_i(O_k) := Pr\{f_X \equiv f_i \mid O_k\}$$

Furthermore, the *vector of beliefs* after k observations is denoted by $[\, B_1(O_k),\ldots,B_m(O_k)\,]$

Notation

After k observations, let f_g be the PDF with the greatest belief. In other words,

$$B_g(O_k) = Max\{B_i(O_k),\, i = 1,\cdots,m\} \tag{1}$$

Similarly, the PDF with the second maximum belief is identified by f_s, or in fact,

$$B_s(O_k) = Max\{B_i(O_k),\, i = 1,\cdots,m \text{ and } i \neq g\} \tag{2}$$

3.1. Updating the Beliefs

After k observations, let assume the decision is to continue. Let x_{k+1} be the outcome of the next observation. Then, $O_{k+1} = (O_k, x_{k+1}) = (x_1,\cdots,x_k,x_{k+1})$ indicates the $(k+1)th$ observations vector. Obviously, the vector of beliefs also changes. To calculate *posterior beliefs* $B_i(O_{k+1})$, $i = 1,\cdots,m$, from the *prior beliefs* $B_i(O_k)$, Bayes formula is applied, as follows:

$$B_i(O_{k+1}) = Pr\{f_X \equiv f_i \mid O_{k+1}\} = \frac{B_i(O_k)\, f_i(x_{k+1})}{\sum_{j=1}^{m} B_j(O_k)\, f_j(x_{k+1})}. \tag{3}$$

4. PAIRWISE SEARCH

As mentioned before, f_i is selected as the desired PDF if it has the greatest belief and this belief is also higher than a predetermined value. For determining this value of least acceptable belief, theoretically we calculate this value for each pair of PDFs and then select the most conservative one. Although the number of PDF pairs is high, it is not necessary to consider all pairs. We show that after each observation actually only one pair is enough to consider.

4.1. Search Space Reduction

Let E be the total decision making space, in which there are m candidates for f_X. Now, we define a subspace $E_{i,j}$ which contains only two candidates, f_i and f_j. In this subspace and after k observations, the belief on f_i is represented by $b_i(O_k)$. Clearly, in $E_{i,j}$,

$$b_i(O_k) + b_j(O_k) = 1 \tag{4}$$

On the other hand, the following relation between beliefs in subspace $E_{i,j}$ and space E is readily seen:

$$b_i(O_k) = \frac{B_i(O_k)}{B_i(O_k) + B_j(O_k)} \tag{5}$$

Lemma 1. *In subspace* $E_{g,i}, \forall i \neq g$, *the value of the belief on* f_g *is higher than* 0.5.

Proof: f_g has the greatest belief from (1). In other words, $b_g(O_k) \geq b_i(O_k)$. Then, $b_g(O_k) \geq 0.5$, from (4).

The first immediate result of this lemma is that the least acceptable belief can be set at least 0.5.

4.2. Reducing Pairwise Comparisons

There are $\binom{m}{2} = m(m-1)/2$ pairwise subspaces. However, since only f_g can be selected, then all subspaces except $E_{g,i}$ are disregarded. Furthermore, we show that $E_{g,s}$ is the only subspace to consider. [f_g and f_s are PDFs with the greatest and second greatest beliefs, respectively by (1) and (2).

To reduce the search to only one pair, we utilize the concept of entropy. Thus, we define entropy, first.

Definition 3. a: Let $\{x_i,\ i = 1, \cdots, n\}$ be the sample space of a discrete random variable X and $p_i = Pr[X = x_i]$. Then, by definition $[-log_2 p_i]$ is called the amount of surprise that one hears x_i is the value of X.

b: Entropy of X is define as follows,

$$H(X) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{6}$$

In information theory, entropy is interpreted as the expected value of uncertainty exists as to the value of X, see Mackay (2005).

Theorem 1. *After* k *observations, let* $B_i(O_k) > B_j(O_k),\ i, j \neq g.$ *Then, the expected value of uncertainty in decision making subspace* $E_{g,i}$ *is higher than that of* $E_{g,j}$.

Proof: See Appendix 1.

Result

From Theorem 1 it is implied that the expected value of highest uncertainty exists in decision making subspace $E_{g,s}$. Therefore, by considering only this subspace, the procedure is stopped in the most conservative case.

5. DETERMINING THE LEAST ACCEPTABLE BELIEF

Let L be the least acceptable belief. The objective is to determine L such that the expected value of correct selection probability be maximized. Obviously, the optimal value of least acceptable belief depends on the number of observations.

5.1. Consistent Grid

In the process of obtaining the optimal value for the least acceptable belief, we need to divide the total domain of belief function $b_g(O_k, x)$ into some non-overlapping intervals such that this function is either increasing or decreasing in each interval. It should be mentioned that $b_g(O_k, x)$ is the belief on f_g in subspace $E_{g,s}$, after k th but before $(k+1)$ th observations, where x is the outcome of a random variable which indicates the next observation.

Consider the plane of $b_g(O_k, x)$ versus x. We create a grid by drawing some horizontal and vertical lines and dividing the plane into some squares.

By Lemma 1, the optimal belief is grater than 0.5. Thus, the range of the belief function starts at 0.5. Let the horizontal lines be drawn at $r_0, r_1, r_2, \cdots, r_\beta$, where $r_0 = 0.5$. Then, the total range is covered by β non-overlapping $R_h = [r_{h-1}, r_h]$, $h = 1,2,\cdots,\beta$. On the other hand, we define D_h, the domain of this function which corresponds with R_h as follows.

$$D_h = \{x : b_g(O_k, x) \in R_h\} \tag{7}$$

Clearly, D_h may consists of some non-overlapping intervals.

Definition 4. The grid created by drawing horizontal and vertical lines on the plane of $b_g(O_k, x)$ versus x is called *Consistent Grid*, if it has the minimum number of lines and within each interval $b_g(O_k, x)$ is either increasing or decreasing, but not both. (If $b_g(O_k, x)$ is less than 0.5, then $[1-b_g(O_k, x)]$, must be either increasing or decreasing, instead.) Furthermore, each domain interval belongs to one R_h interval, only. In Appendix 2, we show how to create the consistent grid.

Example 1

Let us consider $g = 2$, $k = 5$, and $b_2(O_5, x) = \dfrac{x^8}{x^8 + 3337868.212e^{\frac{x}{3}}}$ (details are in the fifth stage of example 2). The extreme points of $b_2(O_5, x)$ are 0 and 24 and the value of $b_2(O_5, x)$ at these points are 0.92 and 0, respectively. Since the value of this function at point 0 is less than 0.5, it is replaced by (1-0) = 1. Therefore, $r_1 = 0.92, r_2 = 1$ and the total range is divided into two intervals, $R_1 = [0.5, 0.92]$ and $R_2 = [0.92, 1]$. Figure 1 shows $b_2(O_5, x)$ versus x and its corresponding consistent grid.

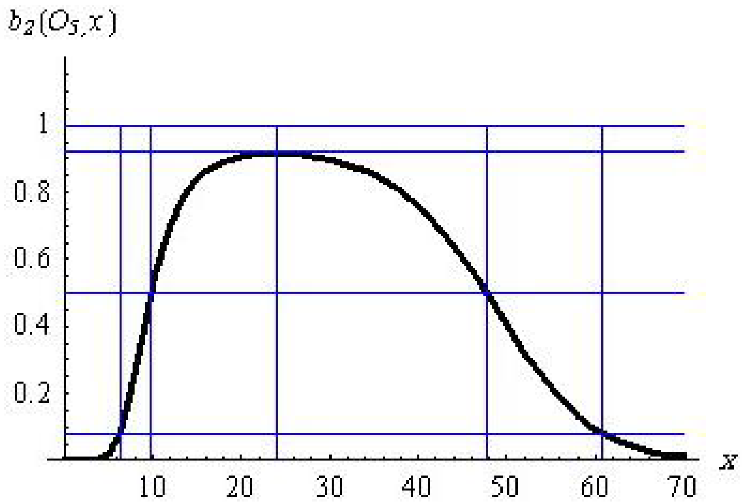

Figure 1. Consistent Grid of $b_2(O_5, x)$ versus x

Here, the value of $b_g(O_k, x)$ at two points, $9.8, 47.7$ is equal to 0.5 and $b_g(O_k, 6.3) = b_g(O_k, 60.7) = 1-0.92$. If the vertical lines are drawn only at 0 and 24, then the resulting grid is not consistent. To see that, consider interval $[0, 24]$. Within $[0, 9.8]$, the value of $b_g(O_k, x)$ is less than 0.5 and $1-b_g(O_k, x)$ is a decreasing function while within $[9.8, 24]$, $b_g(O_k, x)$ is an increasing function. Furthermore, $[0, 9.8]$ corresponds with two different R_h. The first interval, $[0, 6.3] \in R_2$ while $[6.3, 9.8] \in R_1$. Thus, the vertical lines are drawn at

$\{0, 6.3, 9.8, 24, 47.7, 60.7\}$. On the other hand, by (7), $D_1 = \{[6.3, 9.8] \cup [9.8, 24] \cup [24, 47.7] \cup [47.7, 60.7]\}$ and $D_2 = \{[0, 6.3] \cup [60.7, \infty]\}$.

5.2. Stochastic Dynamic Programming Structure

Stochastic dynamic programming is applied to determine the optimal policy. The decision variable is the value of L (the least acceptable belief). Each observation is considered to be one **stage** of dynamic programming. By stage k, we mean k observations have been made, so far. As mentioned before, at most N observations can be made. Thus, at most $N-k$ more observations are allowed at stage k. Let define the following notation in general subspace of $E_{i,j}$.

$U_{i,j}(k)$: the maximum expected value of correct selection probability at stage k in subspace $E_{i,j}$;

As proved in the previous section, after each observation we consider only subspace $E_{g,s}$. Therefore, in this subspace we use the following simplified notation.

$V(k)$: the maximum expected value of correct selection probability at stage k in subspace $E_{g,s}$. In other words, $V(k) = U_{g,s}(k)$.

$V(k:L)$: the maximum expected value of probability of correct selection at stage k in subspace $E_{g,s}$, if L is the least acceptable belief.

$V(k:R_h)$: the maximum expected conditional probability of correct selection at stage k in subspace $E_{g,s}$, given that at next stage the least acceptable belief lies within $R_h = [r_{h-1}, r_h]$.

Then

$$V(k:R_h) = Max\{V(k:L_h), L_h \in R_h\} \tag{8}$$

and

$$V(k) = Max\{V(k:R_h),\ h = 1,\cdots,\beta\} \tag{9}$$

$V(k:L)$ can be calculated by Theorem 2 and then $V(k)$ from (8) and (9).

The Impact of Observation Cost

As mentioned before, a cost is incurred for each observation. To incorporate this cost into the model we assume the real value of the expected probability of correct selection decreases with respect to the number of stages. By considering a discount factor, say $0 < \alpha < 1$, this is accomplished. More precisely, if at stage k, the optimal value the probability of correct selection is $V(k+1)$, then under the same conditions, its value will be $\alpha V(k+1)$, at stage $k+1$.

Theorem 2. *At stage k, if L is the least acceptable belief and α is the discount factor, then,*

$$\begin{aligned} V(k:L) &= [b_g(O_k) - \alpha V(k+1)]Pr[b_g(O_k, X) \geq L] \\ &+ [b_s(O_k) - \alpha V(k+1)]Pr[b_g(O_k, X) \leq 1 - L] + \alpha V(k+1). \end{aligned} \tag{10}$$

Proof: See Appendix 3.

Corollary 1. *For $k \geq 1$,*

$$V(k:1) = \alpha V(k+1), \tag{11}$$

and,

$$V(k) \geq \alpha V(k+1) \tag{12}$$

Proof: If $L = 1$ then the first two terms of the right side of (10) are equal to zero which results in (11). On the other hand, $V(k) \geq V(k:L)$, $\forall L \in [0.5, 1]$ from (9) and (10) or in particular, $V(k) \geq V(k:1)$. Then, (11) implies (12).

Now we consider two special cases where L^* can be determined immediately.

Theorem 3. *In subspace $E_{g,s}$ and at stage k,*

a) Case 1. If $b_g(O_k) \leq \alpha V(k+1)$, then, $L^* = 1$. In other words, the procedure can not be stopped at this stage and $V(k)$ is updated by (11) .

b) Case 2. If $b_s(O_k) \geq \alpha V(k+1)$, then, $L^* = 0.5$. In other words, f_g is selected at this stage.

Proof: By definition of f_g and f_s, $b_s(O_k) < b_g(O_k)$. Then, in case 1 the first two terms of (10) are non-positive and $V(k:L) \leq \alpha V(k+1), \forall L \in [0.5, 1]$. On the other hand, from (12) it is implied that $V(k) = \alpha V(k+1)$. This means continue or in fact, $L^* = 1$.

Similarly, for case 2, it is implied that the first two terms of (10) are non-negative. On the other hand, each probability term of (10) is a nonincreasing function of L. Thus, the maximum value of $V(k)$ is attained at $L^* = 0.5$.

If neither case 1 nor case 2 is satisfied, then,

$$b_s(O_k) < \alpha V(k+1) < b_g(O_k)$$

In this case, we create a consistent grid and optimize the probability of correct selection in each interval separately and then optimize globally.

5.3. Nonlinear Programming Model

Calculating $V(k:L)$ from (10) practically leads to solving a nonlinear programming. In subspace $E_{g,s}$, let a_t and a'_t be the points in the domain, defined as follows.

$$b_g(O_k, a_t) = L, \qquad t = 1,2,\dots,\delta \tag{13}$$

$$b_g(O_k, a'_t) = 1 - L, \qquad t = 1,2,\dots,\delta' \tag{14}$$

If $L \in R_h$, then clearly $a_t, a'_t \in D_h$. On the other hand as mentioned before, since D_h may consist of more than one interval, then for a unique $L \in R_h$, there may be more than one a_t or a'_t within D_h. Therefore, (8) is equivalent to the following.

$$V(k:R_h) = \max_{a \in D_h}\{V(k:L_h), \quad b_g(O_k, a) = L_h, \text{or } b_g(O_k, a) = 1 - L_h\} \tag{15}$$

Therefore, (15) is also equivalent to a nonlinear programming. In this model, the objective function is (10), in which L is substituted with a_t or a'_t, as defined by (13) or (14). Furthermore, the constraints of this nonlinear programming are as follows:

$$b_g(O_k, a_1) = b_g(O_k, a_2) = \cdots = b_g(O_k, a_\delta)$$

$$b_g(O_k, a'_1) = b_g(O_k, a'_2) = \cdots = b_g(O_k, a_{\delta'}')$$

$$b_g(O_k, a_1) + b_g(O_k, a'_1) = 1$$

$$a_t \in D_h, \quad t = 1,\cdots,\delta; \quad a'_t \in D_h, \quad t = 1,\cdots,\delta'.$$

After determining all a_t, and a'_t (the decision variables), the corresponding L is calculated from (13) or (14).

Note

The term $Pr[b_g(O_k, X) \geq L]$ of the objective function of nonlinear programming model (10) can be calculated easily by considering the property of the consistent grid. As mentioned before, the plane of graph $b_g(O_k, x)$ versus x is divided into some nonoverlapping intervals by drawing some vertical and horizontal lines. If this grid is consistent, then $b_g(O_k, x)$ is either increasing or decreasing, but not both at each interval. Based on this property, we calculate $Pr[b_g(O_k, X) \geq L]$ over the individual intervals of the domain. Clearly, if that interval does not belong to D_h, then this term is either 0 or 1, depending on the value of L. However, if this interval, say $[d_{h-1}, d_h] \in D_h$ then,

i) $Pr[b_g(O_k, X) \geq L] = F_X(d_h) - F_X(a_t)$, if $b_g(O_k, x)$ is an increasing function in that interval, where $F_X(.)$ is the CDF of X.

ii) $Pr[b_g(O_k, X) \geq L] = F_X(a_t) - F_X(d_{h-1})$, if $b_g(O_k, x)$ is a decreasing function in that interval. The term $Pr[b_g(O_k, X) \leq 1 - L]$ also can be calculated similarly.

6. ALGORITHM

In this section, we summarize the results of the previous sections and present it as an algorithm.

Initial Step: Start with an initial vector of beliefs and a discount factor α. Determine the decision maker's desired expected value of the probability of correct selection after final experiment, $V(N)$, and set $k := 0$.

Step 1. Set $k := k + 1$ and generate a new observation. Update the vector of beliefs according to (3) and identify f_g and f_s. Determine $b_g(O_k)$ and $b_s(O_k)$ by (5).

Step 2. If $b_g(O_k) \leq \alpha V(k+1)$, then go to Step 1, else If $b_s(O_k) \geq \alpha V(k+1)$, then select f_g as the desired PDF and STOP.

Step 3. Create a consistent grid, by the procedure of Appendix 2.

Step 4. For each range interval $R_h, h = 1, \cdots, \beta,$ solve the nonlinear programming model of Subsection 5.3 and determine $V(k : R_h)$ as well as the corresponding optimum least acceptable belief, L_h^*.

Step 5. Obtain $V(k)$ by (9). Consider $V(k : R_{h^*}) = V(k)$ and set $L^* = L_{h^*}^*$. If $b_g(O_k) \geq L^*$, then select f_g as the desired PDF and STOP, else go to Step 1.

7. NUMERICAL EXAMPLE

Example 2

We apply decision on belief (DOB) technique to determine the life time distribution function of an expensive electronic component. Due to the high price of this component as well as the high cost of experiment, only 10 tests are allowed ($N = 10$). Furthermore, it is desired the distribution function will be identified with the probability of at least 0.95 after 10 experiments, $i.e. V(10) = 0.95$. We also assume $\alpha = 0.98$.
It seems the best distribution function fitting the life time of this component is Gamma with the density function of $f_X(x) = \beta^{-\lambda}(\Gamma(\lambda))^{-1} x^{\lambda-1} e^{-x/\beta}$. Furthermore, the existing information indicates its expected life time is 24 months. Four members of this family of random variables are considered to be the most possible candidate to fit the distribution function with parameters λ and β as follows.

$$\begin{cases} f_1: & Gamma\ (1,24) \\ f_2: & Gamma\ (12,2) \\ f_3: & Gamma\ (4,6) \\ f_4: & Gamma\ (3,8) \end{cases}$$

At the beginning, all four PDFs have equal chance of being selected. Thus,

$$B_1(O_0) = B_2(O_0) = B_3(O_0) = B_4(O_0) = 0.25.$$

Lets generate the random observations from f_2.

First Stage: $k = 1$.

The outcome of the first observation is $x_1 = 35.553$. The vector of belief is updated by (3) as follows.

$$B_1(O_1) = 0.179, B_2(O_1) = 0.253, B_3(O_1) = 0.292, B_4(O_1) = 0.275.$$

Thus, $g = 3, s = 4$. Then, in subspace $E_{3,4}$ and from (5) we have, $b_3(O_1) = 0.515$ and $b_4(O_1) = 0.485$. Since in subspace $E_{34}, b_g(O_1) \le \alpha V(2) = 0.776$, then from case (1) of Theorem 3, $L^* = 1$. Continue and make another observation.

Second Stage: $k = 2$.

The outcome of this observation is $x_2 = 24.298$. Then, after calculation of the belief vector by (3), $g = 2, s = 3$ and in subspace $E_{2,3}, b_2(O_2) = 0.603, b_3(O_2) = 0.397$. Again, from case (1) of Theorem 3, $b_g(O_2) \le \alpha V(3) = 0.893$, it results in $L^* = 1$. Continue and make another observation.

Third Stage: $k = 3$.

The outcome of this observation is $x_3 = 26.464$. Then, $g = 2, s = 3$ and $b_2(O_3) = 0.720, b_3(O_3) = 0.280$. Again, from case (1) of Theorem 3, $L^* = 1$.

Fourth Stage: $k = 4$.

The outcome of this observation is $x_4 = 19.491$. Then, $g = 2, s = 3$ and in subspace $E_{2,3}, b_2(O_4) = 0.793, b_3(O_4) = 0.207$. Again, from case (1) of Theorem 3, $L^* = 1$.

Fifth Stage: $k = 5$.

The outcome of this observation is $x_4 = 20.996$. Then, $g = 2, s = 3$ and in subspace $E_{2,3}, b_2(O_5) = 0.863, b_3(O_5) = 0.137$ and $\alpha V(6) = 0.816$. In this stage we have,

$$b_3(O_5) < \alpha V(6) < b_2(O_5)$$

Thus, we have to solve the nonlinear programming model and determine the optimal value for the least acceptable belief.

The extreme points of $b_2(O_5, x) = \dfrac{x^8}{x^8 + 3337868.212e^{\frac{x}{3}}}$ are 0 and 24, from (16) of Appendix 2.

The only boundary point is 0 (See example 1). Thus, $EP = \{4, 24\}$ and $EP' = \{6.3, 60.7\}$.

We solve the nonlinear programming model for each range interval of R_1 and R_2, separately.
For R_1, let a_1 and a_2 be defined as (13) and (14). Thus,

$$b_2(O_5, a_1) = b_2(O_5, a_2) = L$$

and,

$$b_2(O_5, a_1') = b_2(O_5, a_2') = 1 - L$$

where $a_1 \in [9.8, 24], a_2 \in [24, 47.7], a_1' \in [6.3, 9.8], a_2' \in [47.7, 60.7]$ and,

$$Pr[b_g(O_k, x) \geq L] = F(a_2) - F(a_1)$$

$$Pr[b_g(O_k, x) \leq 1 - L] = [F(a_2) - F(6.3)] + [F(60.7) - F(a_2')]$$

Then, by substituting the terms of (10), the nonlinear programming model is as follows:

$$Max = 0.041[F_2(a_2) - F_2(a_1)] + 0.064[F_3(a_2) - F_3(a_1)] - 0.463[F_2(a_{2'}) - F_2(a_1')]$$

$$-0.93[F_3(a_2') - F_3(a_{1'})] + 1.371$$

Subject to:

$$b_2(O_5, a_1) = b_2(O_5, a_2)$$

$$b_2(O_5, a_1') = b_2(O_5, a_2')$$

$$b_2(O_5, a_1) + b_2(O_5, a_1') = 1$$

$$a_1 \in [9.8, 24], a_2 \in [24, 47.7], a_1' \in [6.3, 9.8], a_2' \in [47.7, 60.7]$$

While, F_2 and F_3 are corresponding CDF of f_2 and f_3, respectively. Solving this model results in,

$$L_1^* = b_2(O_5, a_1^*) = 0.807, V(5:R_1) = 0.855$$

Similarly, the optimal solution for R_2 is found as follows:

$$L_2^* = 1, V(5:R_2) = 0.816.$$

Hence, $V(5) = max\{0.855, 0.816\} = 0.855$ and consequently $L^* = L_1^* = 0.807$. Since $b_2(O_5) \geq L^*$, the procedure stops and selects f_2 as the desired PDF.

Now, suppose that we are going to perform the two celebrated Goodness-of-Fit methods, K-S and Chi-Square, on this problem. Since the Chi-Square method needs at least 50 samples (Conover, 2001), so it is not possible to run it for such small sample of size 5. However, by running the K-S test, all the four candidates are accepted with all p-values of greater than 0.25.

8. CONCLUSION

In this paper, we introduced a new approach for distribution fitting, called *Decision on Beliefs*. This method selects one PDF among a set of candidates to fit the distribution of a random variable X when the number of observations is limited. The basis of this approach is to select the PDF with the highest probability of fitting. However, this probability, called *belief*, is required to be not less than some predetermined value of L, where L depends on the number of observations made. In case the PDF with the greatest belief is less than this value, then another observation is made and the vector of beliefs is updated by Bayesian formula. On the other hand, L is determined by stochastic dynamic programming approach in order to maximize the probability of correct selection. Furthermore, the concept of entropy is also used to reduce the number of necessary comparisons.

There are a vide range numerical examples in Saniee Monfared and Ranaeifar (2007) to evaluate the performance of this new algorithm and compare it with some other celebrated algorithms. DOB has also experienced solving some other statistical problems as *Response Surface methodology* (Eshragh and Akhavan Niaki, 2003) and *Quality Control* (Fallahnezhad *et al*., 2006). In all the cases, adapted DOB algorithms outperform all the best common ones in many aspects such as accuracy of optimal solution, running time and so on.

The chief privilege of DOB is its learning procedure which is revised at each iteration based on collected data and prior beliefs. On the other hand, contrary to the most common algorithms, DOB extracts the disguised information at each generated data and applies them to converge to optimal solution rapidly. It is presumed that this state of the art algorithm can be applied and extended in other areas of statistical analysis, combinatorial optimization problems, and optimization of non-convex programming.

ACKNOWLEDGMENT

We are very grateful to M.A. Saniee Monfared and F. Ranaeifar for proofreading the first manuscript and discussing this problem with us.

APPENDIX

Appendix 1: Proof of Theorem 1

By considering the beliefs of $b_g(O_k)$ and $b_i(O_k)$ in subspace $E_{g,i}$, the value of entropy in this subspace $E_{g,i}$, given the observations vector of O_k, is as follows.

$$H(E_{g,i}) = -b_g(O_k)log_2[b_g(O_k)] - b_i(O_k)log_2[b_i(O_k)] =$$

$$-\frac{B_g(O_k)}{B_i(O_k)+B_g(O_k)}log_2[\frac{B_g(O_k)}{B_i(O_k)+B_g(O_k)}] - \frac{B_i(O_k)}{B_i(O_k)+B_g(O_k)}log_2[\frac{B_i(O_k)}{B_i(O_k)+B_g(O_k)}]$$

The derivative of the entropy results in:

$$\frac{d[H(E_{g,i})]}{d[B_i(O_k)]} = \frac{B_g(O_k)}{[B_i(O_k)+B_g(O_k)]^2}log_2[\frac{B_g(O_k)}{B_i(O_k)}]$$

Since $B_g(O_k) > B_i(O_k), \forall i \neq g$, then this derivative is positive and consequently $H(E_{g,i})$ is an increasing function of $B_i(O_k)$. On the other hand, $B_i(O_k) > B_j(O_k)$ implies that $H(E_{g,i}) > H(E_{g,j})$ or in fact the expected amount of uncertainty in decision making subspace $E_{g,i}$ is higher than that of $E_{g,j}$. Considering the fact that in both subspaces the decision strategy is the same, it is implied that the expected probability of correct selection for f_g in $E_{g,j}$ is higher than for $E_{g,i}$.

Appendix 2: Creating Consistent Grid

In subspace $E_{g,s}$, we create a grid through the following procedure, which is consistent by Definition 4.

Step 1. Identify EP the set of extreme points of $b_g(O_k, x)$ as well as the boundary points. The extreme points are identified by solving the following equation.

$$f_s'(x) f_g(x) = f_s(x) f_g'(x). \quad (16)$$

(This equation results from setting the derivative of $b_g(O_k, x)$ with respect to x equal to zero.)

Step 2. Calculate the value of $b_g(O_k, x_e), x_e \in EP$. If any of these values is less than 0.5, replace it with $1-b_g(O_k, x_e)$. Let $r_1, r_2, \cdots, r_\beta$ be the resulting numbers, after sorting them in ascending order. Now by drawing the horizontal lines at these points, the total range of $[0.5, 1]$ is divided into β intervals of $R_h = [r_{h-1}, r_h]$, $h = 1,2,\cdots,\beta$.

Step 3. Let EP' be the set of points defined as follows:

$$EP' = \{x : b_g(O_k, x) = 1 - b_g(O_k, x_e),\ \ x_e \in EP\}$$

(In fact, the value of $b_g(O_k, x)$ at the points of EP' are the reflect of $r_1, r_2, \cdots, r_\beta$ with respect to 0.5).

The vertical lines are drawn at all points of EP and EP', as well as at the points at which $b_g(O_k, x) = 0.5$. Then, for each range interval of R_h identify D_h, as defined by (7).

Appendix 3: Proof of Theorem 2

At first, we prove this theorem for the general case of subspace $E_{i,j}$, and then it is easy to replace i and j with g and s, respectively. Now, by applying the total probability law we have:

$$\begin{aligned} U_{i,j}(k, L) &= max\{E[Pr(CS \mid O_k, X)]\} = max\{E[Pr(CS \mid S_i, O_k, X)Pr(S_i) \\ &+ Pr(CS \mid S_j, O_k, X)Pr(S_j) + Pr(CS \mid NS_{i,j})Pr(NS_{i,j})]\}. \end{aligned} \tag{17}$$

where events CS, S_i, S_j and $NS_{i,j}$ are defined as follows:

CS : correct selection;
S_i : f_i be selected after the next observation;
S_j : f_j be selected after the next observation;
$NS_{i,j}$: neither f_i nor f_j be selected in the next stage.

To prove Theorem 2, we derive the terms of (17) and substitute them.

$Pr(CS \mid S_i, O_k, X)$ is the probability of correct selection if after the next observation f_i is assumed to be the desired PDF. This is by definition the belief on f_i and is denoted $b_i(O_k, X)$. It is easy to check that $E[b_i(O_k, X)] = b_i(O_k)$

$$Pr(S_i \mid O_k, X) = Pr(\{b_i(O_k, X) > b_j(O_k, X)\} \,\&\, \{b_i(O_k, X) > L\}) = Pr(b_i(O_k, X) > L)$$

The latter equality is because of the result mentioned after Lemma 1, which asserts the least acceptable belief is equal to or greater than 0.5.

$Pr(S_j \mid O_k, X)$ parallels $Pr(S_i \mid O_k, X)$.

$Pr(CS \mid NS_{i,j}) = \alpha U_{i,j}(k+1),$ by definition of $NS_{i,j}$ and discount factor.

$Pr(NS_{i,j}) = 1 - Pr(S_i) - Pr(S_j),$ because $S_i, S_j,$ and $NS_{i,j}$ are exhaustive and mutually exclusive events.

REFERENCES

[1] Ahn J.H., Kim J.J. (1998), Action-Timing Problem with Sequential Bayesian Belief Revision Process; *European Journal of Operational Research* 105; 118-129.

[2] Bernardo J.M., Smith A.F.M. (2001), Bayesian Theory, 1st Edition; Wiley.

[3] Conver W.J. (2001), Practical Nonparametric Statistics, 3rd Edition; Wiley.

[4] Eshragh A. (2001), A New Approach to Distribution Fitting: Decision on Beliefs, *Unpublished B.S. Project*, Industrial Eng. Dept., Sharif University of Technology, Tehran, IRAN.

[5] Eshragh A., Akhavan Niaki S.T. (2003), Application of Decision on Beliefs in Response Surface Methodology; *Proceeding of the 54th Session of International Statistical Institute*; Berlin, Germany.

[6] Fallahnezhad M.S., Akhavan Niaki S.T., Eshragh A. (2006), Application of Decision on Beliefs in Univariate Quality Control Environments; *Proceeding of Operations Research Conference*; Karlsruhe, Germany.

[7] MacKay D.J.C. (2005), Information theory, Inference, and Learning Algorithm; Cambridge University Press; 7th Edition.

[8] Ross S.M. (1983), Introduction to Stochastic Dynamic Programming; Academic Press; New York.

[9] Saniee Monfared M.A., Ranaeifar F. (2007), Further Analysis and Developments of DOB Algorithm on Distribution Fitting; *Scientia Iranica* 14(5); 425-434.