Scientia Iranica, Vol. 14, No. 5, pp 425–434

SCIENTIA IRANICA

Further Analysis and Developments of the Eshragh-Modarres (E-M) Algorithm on Statistical Estimation

M.A.S. Monfared* and F. Ranaiefar[1]

In this paper, the seminal work of Eshragh and Modarres has been discussed in a statistical estimation problem called the Decision on Belief (DoB). The proposed approach has been thoroughly investigated and presented in a novel way, called the 3-phase approach. New instructive examples and detailed calculations are presented to illustrate the logic behind the algorithm in a clear way. The original work has further been developed into new directions, leading to new results.

INTRODUCTION

Ali Eshragh Jahromi and Mohammad Modarres Yazdi [1-3] have developed a new approach for statistical estimation problems called Decision on Beliefs (DoB). In this paper, it is preferred calling the new approach the Eshragh-Modarres algorithm or, simply, the E-M algorithm.

The problem of statistical estimation can be stated in the following way [4]. The random variable, X, with an unknown Probability Distribution Function (PDF), f_X, is given. In order to identify f_X from a set of candidate PDFs, $S = \{f_1, f_2, \cdots, f_m\}$, an algorithm was developed using a special case of an Optimal Stopping Problem [5-8]. At any stage, an experiment is conducted from the presently unknown f_X to generate a new observation and then a decision is made, either to select one of the candidate functions in S, or, to move forward to conduct another experiment. It is assumed that a cost, C, is incurred in obtaining each observation and the total number of possible observations cannot exceed N.

Vector $O_k = (x_1, x_2, \cdots, x_k)$ illustrates the past k observations at stage $i = k$ for $i = 1, 2, \cdots, k, \cdots, N$. Since making a decision is done in a stochastic environment, a probability on the event $\{f_X \equiv f_i\}$ is introduced, i.e., $\Pr\{f_X \equiv f_i\}$, the belief on PDF f_i, which is denoted as $B_i(x_k, O_{k-1})$. By obtaining a new observation, $B_i(x_k, O_{k-1})$ is updated using a formula derived from the Bayes theorem. This formula is used to calculate the posterior beliefs and it is proved that the algorithm is convergent, i.e., after getting enough observations and updating the beliefs with probability one, the belief from which the observations came converges to one and the other beliefs converge to zero.

At any stage, the decision space is confined to $E_{sm,gr}$, representing the subspace containing f_{sm} and f_{gr}, where sm denotes the second best fit candidate for f_x and gr denotes the first best fit candidate for f_x. Note that $B_{gr}(x_k, O_{k-1}) = \max_i\{B_i(x_k, O_{k-1}), i = 1, 2, \cdots m\}$. Within the subspace of $E_{sm,gr}$ and at any stage like k, the strategy for making a decision is: $f_x \equiv f_{gr}$, if $B_{gr}(x_k, O_{k-1}) \geq d_{sm,gr}(n)$ and, otherwise, $k = k + 1$, i.e., a new observation should be taken. The $d_{sm,gr}(n)$, as a real value, defines the expectation of the probability of correct selection and is a threshold for decision making. The value for $d_{sm,gr}(n)$ is calculated using a stochastic dynamic programming approach, in which the expectation of the probability of correct selection is maximized.

In [3], the E-M algorithm has been considered to be much more powerful than the Goodness of Fit techniques, including the Kolmogrov-Smirnov method and the Chi square method. However, it seems that the true strengths of the algorithm lie in the fact that it works in a sequential order and, hence, observations are only generated when needed. This feature is important in applications incurring high cost and risk, such as testing new drugs, prototyping industrial products,

*. *Corresponding Author, Industrial Engineering Group, School of Engineering, Alzahra University, Tehran, I.R. Iran.*
1. *Department of Industrial Engineering, Tarbiat Modarres University, Tehran, I.R. Iran.*

experimenting with nuclear material and launching missiles.

Despite the originality of the work, it has been shown in a recent work [2] that the presentation of the E-M algorithm in its current form is very complicated. In this paper, the algorithm has been presented by a new approach and further developed in new directions. Note that, for proofs and further mathematical analysis, interested readers are referred to [1-3].

The paper is organized as follows. In the following section, the algorithm is systematically presented by a novel 3-phase approach and illustrated using numerical examples. Then, the algorithm is further developed in new directions and the new results are presented. Finally, the paper is concluded and topics for further researches are presented.

E-M ALGORITHM

The primary presentation of the E-M algorithm is very complicated [1-2], where one can hardly follow the logic behind it. In this section, a systematic approach is developed illustrating the working logic of the E-M algorithm in a novel way. The steps required to solve a problem have been broken into a 3-phase procedure emphasizing working logic rather than mathematical proof. Note that the algorithm becomes more sophisticated and more effective as it moves from Phase 1 to Phase 3. However, in the authors' presentation, one may stop at the end of Phase 1 (or Phase 2) and completely have a solution, which is presently a formidable task. In this case, however, a larger number of observations may be needed. Also, numerical examples and graphical illustrations have been presented, enhancing the understanding of the algorithm.

Phase One (Preliminaries)

Step i

Define $S = \{f_1, f_2, \cdots f_m\}$, i.e., the set of candidate probability functions, where all m functions have been considered appropriate, primarily for f_x, the unknown best fit probability function.

Step ii

Initialize $B_i() = \frac{1}{m}$, as the prior belief value for the ith candidate, considering the maximum entropy principle. Also, set α as the discount rate, $V(N)$ as the maximum probability of correct selection and N as the maximum number of observations which can be generated in the experiment.

Step iii

Set $k = 0$.

Step iv

Conduct an experiment to generate x_k from f_x.

Step v

Estimate the posterior belief values, $B_i()$ (for $i = 1, 2, \cdots m$), by using the following:

$$B_i(O_k) = B_i(x_k, O_{k-1}) = \frac{B_i(O_{k-1}).f_i(x_k)}{\sum_{j=1}^{m} B_j(O_{k-1}).f_j(x_k)}.$$

Step vi

Build order statistics on posterior beliefs, $B_i()$ as $B_{(1)} < B_{(2)} < \cdots < B_{(m-1)} < B_{(m)}$, where (m) denotes the greatest belief and (1) the least belief obtained, respectively. In other words, $B_{(m)}() = \max\{B_1(), B_2(), \cdots, B_m()\}$. For the sake of brevity, $B_{(m-1)}()$ and $B_{(m)}()$ are denoted as $B_{sm}()$ and $B_{gr}()$, respectively.

Step vii

Normalize $B_{sm}()$ and $B_{gr}()$ using the following:

$$B_{sm,gr}(sm; O_k) = \frac{B_{sm}(O_k)}{B_{sm}(O_k) + B_{gr}(O_k)},$$

and:

$$B_{sm,gr}(gr; O_k) = \frac{B_{gr}(O_k)}{B_{sm}(O_k) + B_{gr}(O_k)}.$$

Note that $B_{sm,gr}(sm; O_k) + B_{sm,gr}(gr; O_k) = 1$. These steps are further illustrated in the following example.

Example 1

Consider $S = \{f_1, f_2, f_3, f_4\}$, where f_1 = Gamma $(3, 4)$, f_2 = Gamma $(12, 2)$, f_3 = Gamma $(16, \sqrt{3})$, f_4 = Gamma $(4, 2\sqrt{3})$ and $B_i() = \frac{1}{4}$ = 0.25. Random numbers have been generated for f_2 using Minitab and Steps i to v have been implemented. Results are shown in Figure 1. As seen from Figure 1, $B2$ approaches 1 around $k = 40$ illustrating that f_2 is the winner function.

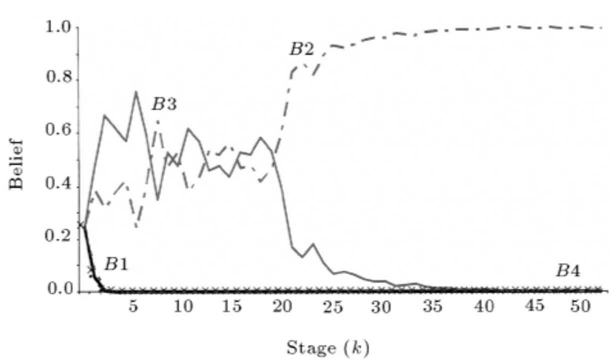

Figure 1. Converging trends of four different belief functions (Example 1).

Phase Two (Correct Selection)

Step viii

If $B_{sm,gr}(sm;O_k) > \alpha.V^*_{sm,gr}(k+1)$, then, $f_x = f_{gr}$ is the best fit function and one should terminate. Note that $B_{sm,gr}(sm;O_k)$ can be denoted in a simpler form of either $B(sm;O_k)$ or $B_{sm}(O_k)$. Similarly, $B_{sm,gr}(gr;O_k)$ can be denoted as either $B(gr;O_k)$ or $B_{gr}(O_k)$. Also, note that $\alpha.V^*_{sm,gr}(k+1)$ can be denoted as $\alpha.V(k+1)$ for the sake of brevity and is determined by $\alpha.V(k+1) = \alpha^{N\;\;k} * V(N)$.

Step ix

If $B_{sm,gr}(gr;O_k) < \alpha.V^*_{sm,gr}(k+1)$, then, $f_x \neq f_{gr}$, so that taking a new observation is required, i.e., if $K \leq N$ set $k = k+1$, then, go to Step iv; otherwise (i.e., if $k > N$) stop, then, $f_x \equiv f_{gr}$ is the best fit function and one should terminate.

Step x

If $B_{sm,gr}(sm;O_k) < \alpha.V^*_{sm,gr}(k+1) < B_{sm,gr}(gr;O_k)$ and $B_{sm,gr}(gr;O_k) \geq d^*(k)$, then, $f_x = f_{gr}$, else, generate a new observation, i.e., if $K \leq N$ set $k = k+1$, then, go to Step iv; otherwise (i.e., if $k > N$) stop, then, $f_x = f_{gr}$ is the best fit function. Note that $d^*(k)$ is estimated according to a procedure developed in the following section. The complete decision making procedure is also shown in Figure 2.

Phase Three (Estimating $d^*(k)$)

Consider $d^*(k)$ as a decision making criteria or a threshold by which the best fit function can be determined efficiently. The procedure to determine $d^*(k)$ is considered in the following steps.

Step xi

Define $y_{x_{k+1}}$ as the most plausible belief on f_{gr} as:

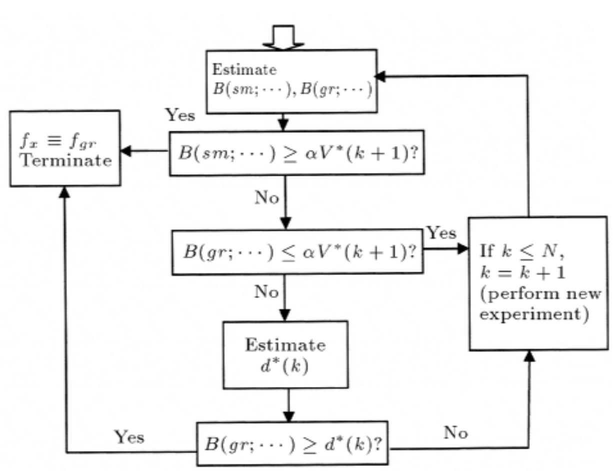

Figure 2. Decision making procedure.

$$y_{x_{k+1}} = B_{sm,gr}(gr;x_{k+1},O_k) =$$

$$\frac{B_{sm,gr}(gr;O_k).f_{gr}(x_{k+1})}{B_{sm,gr}(gr;O_k).f_{gr}(x_{k+1}) + B_{sm,gr}(sm;O_k).f_{sm}(x_{k+1})}.$$

Note that x_{k+1} has not been generated yet and it is assumed that it is the best possible observation one can expect to have at the present stage to select f_{gr} as f_x. Under this assumption, one considers estimating the highest plausible belief one can get on the present best fit function, f_{gr}. The underlying idea here is that, if the next forthcoming observation were considered to be the best possible one, would it be possible to terminate the process and make a decision on a best fit function or not? This idea, as illustrated in the following, will help to minimize the need for additional experiments.

Example 2

Suppose that $B_{sm,gr}(sm;O_8) = 0.471$, $B_{sm,gr}(gr;O_8) = 0.529$, $f_{sm}(x) = \frac{1}{\pi(1+x^2)}$, $f_{gr}(x) = \frac{1}{\sqrt{2\pi}} e^{\;\frac{x^2}{2}}$ and, hence:

$$y(x_{k+1})$$

$$= \frac{B_{sm,gr}(gr;O_k).f_{gr}(x_9)}{B_{sm,gr}(gr;O_k).f_{gr}(x_9) + B_{sm,gr}(sm;O_k).f_{sm}(x_9)},$$

or:

$$y(x_9) = \frac{0.529 * \frac{1}{\sqrt{2\pi}} e^{\;\frac{x_9^2}{2}}}{0.529 * \frac{1}{\sqrt{2\pi}} e^{\;\frac{x_9^2}{2}} + 0.471 * \frac{1}{\pi(1+x_9^2)}},$$

as illustrated in Figure 3. Note that x_9 has not yet been realized by experimentation, but it is known that its value could only change to the extent shown in Figure 3.

Step xii

Find a derivative of $y_{x_{k+1}}$, with respect to x_{k+1}, and set this equal to 0, i.e., $f'_{sm}(x_{k+1}).f_{gr}(x_{k+1}) = f_{sm}(x_{k+1}).f'_{gr}(x_{k+1})$, to obtain the roots of the equation, i.e., $x_{k+1,t}$, for $t = 0, 1, \ldots l$.

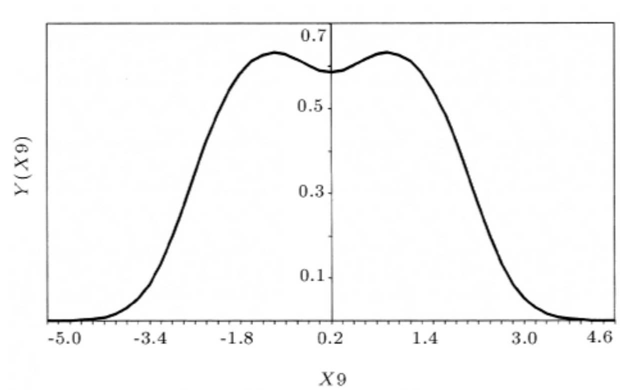

Figure 3. $y(x_9)$ versus x_9 (Example 2).

Step xiii

Find $y_{x_{k+1}}$'s for corresponding $x_{k+1,t}$ which are denoted as $y_0, y_1, \cdots, y_l$ (or y_t for $t = 0, 1, \cdots l$).

Step xiv

Define the reflect lines, y_{Re_t}, with respect to $y = 0.5$, as $y_{\text{Re}_t} = 1 \quad y_t$ for $t = 0, 1, \cdots l$. Here, at most, $2(l+1)$ distinct lines can be drawn.

Step xv

Cross y_t and y_{Re_t} lines with a $y_{x_{k+1}}$ curve and find the corresponding points on $x_{k+1,t}$; Hence, the $x_{k+1,t}$ line is divided into segments which are denoted as $I_1, I_2, \cdots I_{\eta+1}$. Also, the $y_{x_{k+1}}$ line is divided into segments which are denoted as $J_1, J_2, \cdots J_s$. Also, J_s is denoted as J_{Re_s}, if $J_s < 0.5$. Note that, for any J_s segment, there could be more than one I_t segment, where $t = 1, 2, \cdots, \eta + 1$. In the following example, these steps are further illustrated.

Example 3

By setting $y'(x_9) = 0$, one has $x_{9,1} = 0, x_{9,2} = 1, x_{9,3} = 1$ and the corresponding values for y_{x_9} will be $y_0 = 0$, $y_1 = 0.585$, $y_2 = 0.631$ and $y_3 = 0.631$. The associated reflect lines, with respect to y_t's, are illustrated in Table 1. The values for J_ss and J_{Re_s}s are also illustrated in Table 2.

Note that, for both $y = 0$ and $y = 1$, there are 7 lines, which, when crossed by $y(x_9) = B_{1,3}(3; x_9, O_8)$, will have the following 11 points ($\eta = 11$):

Table 1. Reflect lines y_t and its associated lines y_{Re_t}.

y_t	y_{Re_t}
$y_0 = 0$	$y_{\text{Re}_0} = 1$
$y_1 = 0.585$	$y_{\text{Re}_1} = 0.415$
$y_2 = 0.631$	$y_{\text{Re}_2} = 0.369$
$y_3 = 0.631$	$y_{\text{Re}_3} = 0.369$

Table 2. Values for J_ss and J_{Re_s}.

s	J_s	J_{Re_s}
1	$J_1 \equiv [0.5, 0.585]$	$J_{\text{Re}_1} \equiv [0.415, 0.5]$
2	$J_2 \equiv [0.585, 0.631]$	$J_{\text{Re}_2} \equiv [0.369, 0.415]$
3	$J_3 \equiv [0.631, 1]$	$J_{\text{Re}_3} \equiv [0, 0.369]$

$xc_{9,1} = \quad 2.350, \qquad xc_{9,2} = \quad 2.225,$

$xc_{9,3} = \quad 1.963, \qquad xc_{9,4} = \quad 1.586,$

$xc_{9,5} = \quad 1, \qquad xc_{9,6} = 0,$

$xc_{9,7} = 1, \qquad xc_{9,8} = 1.586,$

$xc_{9,9} = 1.963, \qquad xc_{9,10} = 2.225,$

$xc_{9,11} = 2.350.$

Since $y(x_9) = B_{1,3}(3; x_9, O_8)$ is an even function, then, one has the symmetric roots of $xc_{9,t} = xc_{9,12 \quad t}$; $t = 1, \cdots, \frac{\eta+1}{2} = 6$. Now, due to the fact that $\eta + 1 = 12$, one needs to divide the x_{k+1}- axis into 12 segments, as illustrated in Table 3 and shown, also, in Figure 4.

Table 3. Values of segments at ions.

$I_1 \equiv (\quad \infty, \quad 2.350]$	$I_2 \equiv (\quad 2.350, \quad 2.225]$
$I_3 \equiv (\quad 2.225, \quad 1.963]$	$I_4 \equiv (\quad 1.963, \quad 1.586]$
$I_5 \equiv [\quad 1.586, \quad 1]$	$I_6 \equiv [\quad 1, 0]$
$I_7 \equiv [0, 1]$	$I_8 \equiv [1, 1.586]$
$I_9 \equiv [1.586, 1.963]$	$I_{10} \equiv [1.963, 2.225]$
$I_{11} \equiv [2.225, 2.350]$	$I_{\eta+1=12} \equiv [2.350, +\infty]$

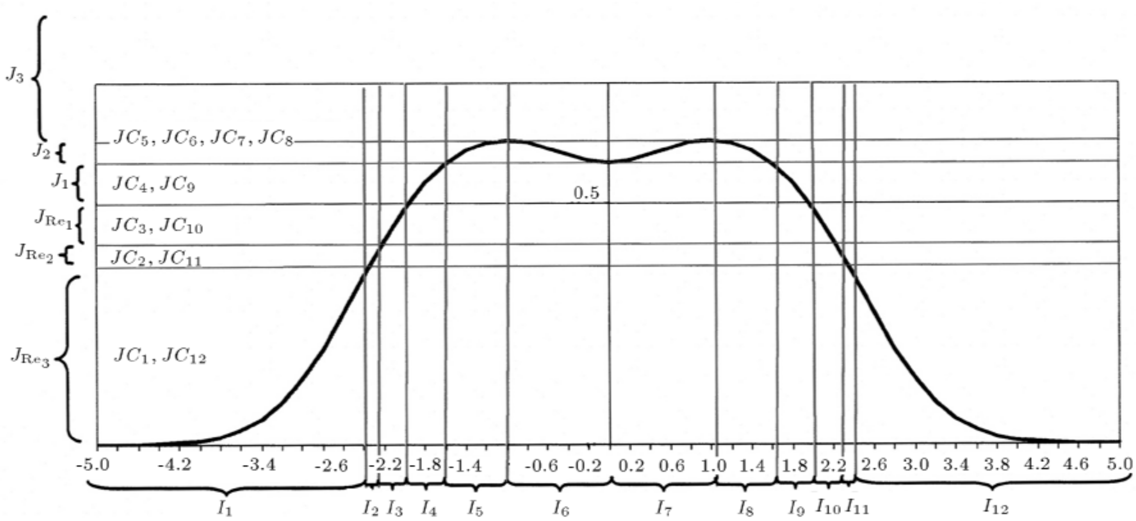

Figure 4. I_t segments versus JC_t segments (Example 3).

Step xvi

For any I_t segment on the x_{k+1}- axis, define a corresponding JC_t segment on the $y(x_{k+1})$-axis. This produces $\eta + 1$ sub functions, i.e., for any $x_{k+1} \in I_t$ and $y(x_{k+1}) \in JC_t$.

Step xvii

Collect the monotonically increasing segments in one set, S_{In}, and the monotonically decreasing segments in another set, S_{De}.

Step xviii

Calculate ψ_t for $t = 1, 2, \cdots, \eta + 1$, using:

$$\Psi_t = \Pr\{x_{k+1} \in I_t\}$$

$$= \int\limits_{I_t} (B_{sm,gr}(sm; O_k).f_{sm}(x)$$

$$+ B_{sm,gr}(gr; O_k).f_{gr}(x))dx$$

$$= B_{sm,gr}(sm; O_k). \int\limits_{I_t} f_{sm}(x)dx$$

$$+ B_{sm,gr}(gr; O_k). \int\limits_{I_t} f_{gr}(x)dx.$$

Note that the function $t = 1, 2, \cdots, \frac{\eta+1}{2}$, if $y_{x_{k+1}}$ is an even function. This is further illustrated in the following example.

Example 4

Since, for any I_t segment in the x_9- axis, one has a corresponding JC_t segment on the $y(x_9)$-axis, it can be seen from Figure 4 that:

$$JC_1 \equiv J_{\mathrm{Re}_3}, \qquad JC_2 \equiv J_{\mathrm{Re}_2}, \qquad JC_3 \equiv J_{\mathrm{Re}_1},$$

$$JC_4 \equiv J_1, \qquad JC_5 \equiv J_2, \qquad JC_6 \equiv J_2$$

$$JC_7 \equiv J_2, \qquad JC_8 \equiv J_2, \qquad JC_9 \equiv J_1,$$

$$JC_{10} \equiv J_{\mathrm{Re}_1}, \quad JC_{11} \equiv J_{\mathrm{Re}_2}, \quad JC_{12} \equiv J_{\mathrm{Re}_3}.$$

Now, it is clear that I_1, I_2, I_3, I_4, I_5 and I_7 form the monotonically increasing set, $S_{In} = \{I_t, \forall t = 1, 2, 3, 4, 5, 7\}$ and that I_6, I_8, I_9, I_{10}, I_{11} and I_{12} form the monotonically decreasing set, $S_{De} = \{I_t, \forall t = 6, 8, 9, 10, 11, 12\}$.

Now, one has to calculate the probabilities, ψ_t, as follows:

$$\psi_t = B_{1,3}(1; O_8). \int\limits_{I_t} f_1(x)dx + B_{1,3}(3; O_8). \int\limits_{I_t} f_3(x)dx$$

$$= 0.471 \int\limits_{I_t} \frac{1}{\pi(1 + x_9^2)} dx + 0.529 \int\limits_{I_t} \frac{1}{\sqrt{2\pi}} e^{\quad \frac{x_9^2}{2}} dx.$$

Hence:

$$\Psi_1 = \Psi_{12} = 0.0652, \qquad \Psi_2 = \Psi_{11} = 0.0049,$$

$$\Psi_3 = \Psi_{10} = 0.0269, \qquad \Psi_4 = \Psi_9 = 0.0168,$$

$$\Psi_5 = \Psi_8 = 0.2476, \qquad \Psi_6 = \Psi_7 = 0.1386.$$

Note that $y = B_{1,3}(3; x_9, O_8)$ is an even function, so that only six sub functions (i.e., $\frac{\eta+1}{2}$) need to be considered.

Step xix

Draw two lines of $y = d_t(k)$ and $y = 1 \quad d_t(k)$ and cross these two lines with $y(x_{k+1}) = B_{sm,gr}(gr; x_{k+1}, O_k)$ to produce the following two corresponding points on the x-axis, a_t and b_t. Repeat this for $t = 1, 2, \cdots, \eta + 1$. See the following example for better illustrations.

Example 5

To estimate $d^*_{s=1}(k)$ for $t = 1$, for example, two lines of $y = d_1(k)$ and $y = 1 \quad d_1(k)$ should be drawn, as illustrated in Figure 5. Cross these lines with $y(x_{k+1}) = B_{sm,gr}(gr; x_{k+1}, O_k)$ and produce two points on the x-axis, denoted as a_1 and b_1.

Step xx

Construct the following dynamic programming model to maximize the probability of correct selection in the sth segment with k observations (simpler notations have been adopted by setting $d^*(J_s, k+1)$ and $\max_{d(J_s;n) \in J_s}$ by $d^*_s(k)$ and $\max_{d_s(k)}$, respectively. Also, V_{sm} and V_{gr} have been introduced to simplify the presentation of the equation. The dynamic model is,

$$V^*_s(k) = \max_{d_t(k)} \{V_{gr} + V_{sm} + \alpha.V^*(k+1)\},$$

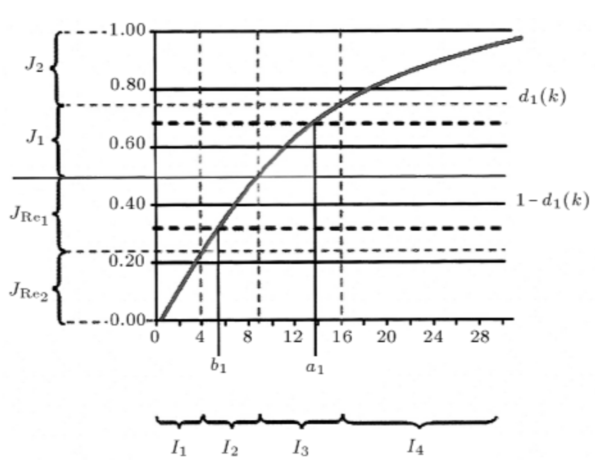

Figure 5. Illustrating an example for $y = d_1(k)$ and $y = 1 \quad d_1(k)$ (Example 5).

where:

$$V_{gr} = [B_{sm,gr}(gr; O_k) \quad \alpha.V_{sm,gr}(d_t^*(k+1))]$$

$$* \left[\sum_t \Pr(B_{sm,gr}(gr; x_{k+1}, O_k) \geq d_t(k) | x_{k+1} \in I_t).\Psi_t \right],$$

$$V_{sm} = [B_{sm,gr}(sm; O_k) \quad \alpha.V_{sm,gr}(d_t^*(k+1)]$$

$$* \left[\sum_t \Pr(B_{sm,gr}(gr; x_{k+1}, O_k) \leq 1 \quad d_t(k) | x_{k+1} \in I_t).\Psi_t \right].$$

Step xxi

Calculate the ingredient probabilities, Pr(..) of V_{gr} and V_{sm}, as follows:

$$\Pr(B_{sm,gr}(gr; x_{k+1}, O_k) \geq d_t(k) | x_{k+1} \in I_t)$$

$$= \begin{cases} 0, & \text{If } JC_t < J_s \\ \Pr\{x_{k+1} \geq a_t | x_{k+1} \in I_t\}, & \text{If } JC_t \equiv J_s \\ & \text{and } t \in S_{In} \\ \Pr\{x_{k+1} \leq a_t | x_{k+1} \in I_t\}, & \text{If } JC_t \equiv J_s \\ & \text{and } t \in S_{De} \\ 1, & \text{If } JC_t > J_s \end{cases}$$

and:

$$\Pr(B_{sm,gr}(gr; x_{k+1}, O_k) \leq 1 \quad d_t(k) | x_{k+1} \in I_t)$$

$$= \begin{cases} 1, & \text{If } JC_t < J_{\mathrm{Re}_s} \\ \Pr\{x_{k+1} \leq b_t | x_{k+1} \in I_t\}, & \text{If } JC_t \equiv J_{\mathrm{Re}_s} \\ & \text{and } t \in S_{In} \\ \Pr\{x_{k+1} \geq b_t | x_{k+1} \in I_t\}, & \text{If } JC_t \equiv J_{\mathrm{Re}_s} \\ & \text{and } t \in S_{De} \\ 0, & \text{If } JC_t > J_{\mathrm{Re}_s} \end{cases}$$

Note that:

$$\begin{cases} \Pr\{x_{k+1} \leq x | x_{k+1} \in I_t\} = \frac{F_{x_{k+1}}(x) \quad F_{x_{k+1}}(xc_{k+1,t \quad 1})}{\Psi_t} \\ \Pr\{x_{k+1} \geq x | x_{k+1} \in I_t\} = \frac{F_{X_{k+1}}(xc_{k+1,t}) \quad F_{x_{k+1}}(x)}{\Psi_t} \end{cases}.$$

where,

$$F_{X_{k+1}}(x) = B_{sm,gr}(sm; O_k).F_{sm}(x)$$

$$+ B_{sm,gr}(gr; O_k).F_{gr}(x),$$

where F denotes cummulative distribution function.

Step xxii

Formulate the following nonlinear dynamic programming model to solve $d_s^*(k)$ for each sub problem, J_s,

$$V_s^*(k) = \max_{d_t(k)} \{V_{gr} + V_{sm} + \alpha.V^*(k+1)\}.$$

Subject to:

$$B_{sm,gr}(gr; a_1, O_k) = B_{sm,gr}(gr; a_2, O_k) = \cdots$$

$$= B_{sm,gr}(gr; a_\alpha, O_k),$$

$$B_{sm,gr}(gr; b_1, O_k) = B_{sm,gr}(gr; b_2, O_k) = \cdots$$

$$= B_{sm,gr}(gr; b_\beta, O_k),$$

$$B_{sm,gr}(gr; a_1, O_k) + B_{sm,gr}(gr; b_1, O_k) = 1,$$

$$a_l \in I_l, \text{ for } l = 1, 2, \cdots, \alpha,$$

$$b_l \in I_l, \text{ for } l = 1, 2, \cdots, \beta.$$

Here, again, both V_{gr} and V_{sm} are defined as in Step xx. Also, the last two constraints can be associated with $d_s(k) \in J_s$. Both α and β define the number of intervals that $y(x_{k+1})$ changes within J_s and J_{Re_s}, respectively. Note that the first three constraints can also be stated in the following forms,

$$y_{x_{k+1}}(a_1) = y_{x_{k+1}}(a_2) = \cdots = y_{x_{k+1}}(a_\alpha),$$

$$y_{x_{k+1}}(b_1) = y_{x_{k+1}}(b_2) = \cdots = y_{x_{k+1}}(b_\alpha),$$

$$y_{x_{k+1}}(a_1) + y_{x_{k+1}}(b_1) = 1.$$

Example 6

To estimate $V^*(5)$, one needs to model and solve $V_1^*(5)$, $V_2^*(5)$ and $V_3^*(5)$, each for an interval shown in Table 4.

Let one now solve $V_1^*(5)$ as:

$$s = 1, \qquad k = 5, \qquad \alpha.V^*(k+1) = 0.51,$$

$$B_{sm,gr}(sm; x_9, O_8) = B_{1,3}(1; x_9, O_8) = 0.4705,$$

$$B_{sm,gr}(gr; x_9, O_8) = B_{1,3}(3; x_9, O_8) = 0.5294,$$

then,

$$V_1^*(5) = \max_{J_1} \Big\{ (0.470588236 \quad 0.51)(1.\Psi_1 + 1.\Psi_2$$

$$+ \Pr\Big\{ B_{1,3}(3; x_9, O_8) \leq 1 \quad d_1(5) | x_9 \in I_3 \Big\}.\Psi_3$$

$$+ 0.\Psi_4 + 0.\Psi_5 + 0.\Psi_6 + 0.\Psi_7 + 0.\Psi_8 + 0.\Psi_9$$

Table 4. Division of domains for $V^*(5)$.

$V_1^*(5)$	$J_1 \equiv$ [0.5, 0.585058521]
$V_2^*(5)$	$J_2 \equiv$ [0.585058521, 0.631049441]
$V_3^*(5)$	$J_3 \equiv$ [0.631049441, 1]

$$+\Pr\left\{B_{1,3}(3;x_9,O_8)\leq 1 \quad d_1(5)|x_9\in I_{10}\right\}$$

$$.\Psi_{10}+1.\Psi_{11}+1.\Psi_{12}$$

$$+(0.529411764 \quad 0.51)(0.\Psi_1+0.\Psi_2+0.\Psi_3$$

$$+\Pr\left\{B_{1,3}(3;x_9,O_8)\geq d_1(5)\middle|x_9\in I_4\right\}.\Psi_4$$

$$+1.\Psi_5+1.\Psi_6 0+.\Psi_7+1.\Psi_8+0.\Psi_8$$

$$+\Pr\left\{B_{1,3}(3;x_9,O_8)\leq 1 \quad d_1(5)|x_9\in I_9\right\}$$

$$.\Psi_9+0.\Psi_{10}+0.\Psi_{11}+0.\Psi_{12})+0.51\Bigg\},$$

or:

$$V_1^*(5)=\max_{J_1}\left\{(\quad 0.039411764).\left(\sum_{t=1}^{2}\Psi_t\right.\right.$$

$$+\Pr\{B_{1,3}(3;x_9,O_8)\leq 1 \quad d_1(5)|x_9\in I_3\}.\Psi_3$$

$$+\Pr\{B_{1,3}(3;x_9,O_8)\leq 1 \quad d_1(5)|x_9\in I_{10}\}.\Psi_{10}$$

$$\left.+\sum_{t=11}^{12}\Psi_t\right)+(0.019411764).(\Pr\{B_{1,3}(3;x_9,O_8)$$

$$\geq d_1(5)\Big|x_9\in I_4\}.\Psi_4+\sum_{t=5}^{8}\Psi_t+\Pr\{B_{1,3}(3;x_9,O_8)$$

$$\geq d_1(5)\Big|x_9\in I_9\}.\Psi_9)+0.51.$$

Since, $y=B_{1,3}(3;x_9,O_8)$ is an even function, one will have:

$$\begin{cases}\Pr\{B_{1,3}(3;x_9,O_8)\leq 1 \quad d(1;5)|x_9\in I_3\}\\ \quad=\Pr\{B_{1,3}(3;x_9,O_8)\leq 1 \quad d_1(5)|x_9\in I_{10}\}\\ \Pr\{B_{1,3}(3;x_9,O_8)\geq d_1(5)|x_9\in I_4\}\\ \quad=\Pr\{B_{1,3}(3;x_9,O_8)\geq d_1(5)|x_9\in I_9\}\end{cases},$$

so that, $V_1^*(5)$ can be rewritten as:

$$V_1^*(5)=\max_{J_1}\left\{(\quad 0.078823528).(\Pr\{B_{1,3}(3;x_9,O_8)\right.$$

$$\leq 1 \quad d_1(5)|x_9\in I_3\}.\Psi_3)+(0.038823528)$$

$$.(\Pr\{B_{1,3}(3;x_9,O_8)\geq d_1(5)|x_9\in I_4\}$$

$$\left..\Psi_4)+0.519468117\right\}.$$

The nonlinear programming model to solve is then,

$$V_1^*(5)=\max\{(\quad 0.078823528).(F_{X_9}(b) \quad F_{X_9}(\quad 2.225))$$

$$+(0.038823528).(F_{X_9}(\quad 1.586) \quad F_{X_9}(a))$$

$$+0.519468117\}.$$

Subject to:

$$(0.8).\left(\frac{f_1(a)}{f_3(a)}\right)=(1.125).\left(\frac{f_3(b)}{f_1(b)}\right),$$

$$1.963\leq a\leq \quad 1.586, \qquad 2.225\leq b\leq \quad 1.963.$$

Now, this problem is solved by writing the program in Lingo [9], as shown in Figure 6.

In writing the Lingo program, shown in Figure 6, the following notes can be helpful:

1. Since the standard normal probability function in Lingo, denoted as @psn(x), can only accept positive values, the negative values have been transformed into positive ones by using $\Phi(\quad x)=1 \quad \Phi(x)$. This is correct, due to the symmetric nature of the normal distribution function;

2. Also, since the Cauchy probability function has not been defined in Lingo, the t-student probability function is used, with one degree of freedom, denoted as @ptd(n,x) in Lingo, to produce almost the same results. In this case, the $t_1(\quad x)=1 \quad t_1(x)$ transformation is used to produce positive values from negative ones, as the t-student function is also a symmetric function;

```
model:
max = −0.078823528 ∗ (0.470588236 ∗ (1− @ptd(1, b))
+0.529411764 ∗ (1 − @psn(b))
−0.470588236 ∗ (1 − @ptd(1, 2.225))
−0.529411764 ∗ (1 − @psn(2.225)))
+0.038823528 ∗ (0.470588236 ∗ (1 − @ptd(1, 1.586))
+0.529411764 ∗ (1 − @psn(1.586))
−0.470588236 ∗ (1 − @ptd(1, a))
−0.529411764 ∗ (1 − @psn(a))) + 0.519468117;
0.503 ∗ @ exp((a^2 + b^2)/2) = (1 + b^2) ∗ (1 + a^2);
1.963 >= a;
a >= 1.586;
2.225 >= b;
b >= 1.963;
d = 1/(1 + 0.709230722 ∗ @ exp(a^2/2)/(1 + a^2));
End
```

Figure 6. Lingo program.

Table 5. Finding the optimal decision value.

s	J_s	$V_s^*(k)$	$d_s^*(k)$
1	J_1	$V_1^*(5) = 0.5197342$	$d_1^*(5) = 0.5402861$
2	J_2	$V_2^*(5) = 0.5194529$	$d_2^*(5) = 0.5850582$
3	J_3	$V_3^*(5) = 0.51$	$d_3^*(5) = 1.0$

The solutions to the non-linear program are:

$$a = 1.806395,$$

$$b = 2.091559 \Rightarrow V_1^*(5) = 0.5197342,$$

$$d_1^*(5) = B_{1,3}(3; a^*, O_8) = 1 \quad B_{1,3}(3; b^*, O_8)$$

$$= 0.5404193.$$

Repeating the above procedure for $V_2^*(5)$ and $V_3^*(5)$ leads to the results illustrated in Table 5 (see [2,3] for details).

Hence, $V^*(k) = \max_{s=1,2,3}\{V_s^*(k)\} = 0.5197342$, then $d^*(k) = d_1^*(5) = 0.5402861$. This completes the numerical illustrations.

FURTHER DEVELOPMENTS

In this section, reports are made on further developments of the E-M algorithm. First, parameter optimization of N is considered and, then, a case is considered where both continuous and discrete functions can be evaluated in a mixed format.

Parameter Optimization, N

Parameter optimization is an important element for any efficient algorithm [10] including the E-M algorithm. Currently, in the E-M algorithm, the parameter N, i.e. the number of observations, has not received adequate attention and it is not clear how it could be estimated. If N is not large enough, then it will not be possible to guarantee the convergence of the algorithm. In other words, small N may lead to the wrong selection of the candidate function. The question is: How can the value of N be estimated? This is the subject of this Section. Let one start with the following two cases, illustrated in Examples 2 and 3.

Example 7

Consider the following case, where f_4 is the true candidate function,

$$f_1 = \text{Normal }(25, 100), \qquad f_2 = \text{Normal }(21, 100),$$

$$f_3 = \text{Normal }(23, 100), \qquad *f_4 = \text{Normal }(22, 100),$$

$$f_5 = \text{Normal }(24, 100), \qquad f_6 = \text{Normal }(20, 100).$$

To simulate this case, 10,000 random numbers have been used, generated by Minitab. The result as illustrated in Figure 7, shows that at least 7000 observations are needed to enable f_4 to converge. Hence, in this case, the parameter, N, should be set around 7000, which is an extremely large number. Intuitively, it can be seen that the large variance, associated with the candidate function, could be a reason.

Example 8

Consider, again, the following case, where f_4 is the true candidate function,

$$f_1 = \text{Normal }(25, 4), \qquad f_2 = \text{Normal }(21, 4),$$

$$f_3 = \text{Normal }(23, 14), \qquad *f_4 = \text{Normal }(22, 4),$$

$$f_5 = \text{Normal }(24, 4), \qquad f_6 = \text{Normal }(20, 4).$$

The difference here, with respect to Example 7, is the smaller variance of the corresponding functions. Figure 8 illustrates the converging process. Here, the proper N is about 60 observations, which are drastically smaller than the 7000 observations required in the previous case.

Now that the importance of the right selection of N has been shown, a new procedure is proposed

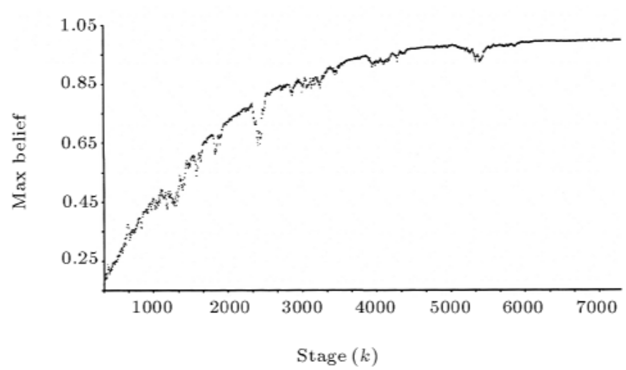

Figure 7. Simulation of f_4 (Example 7).

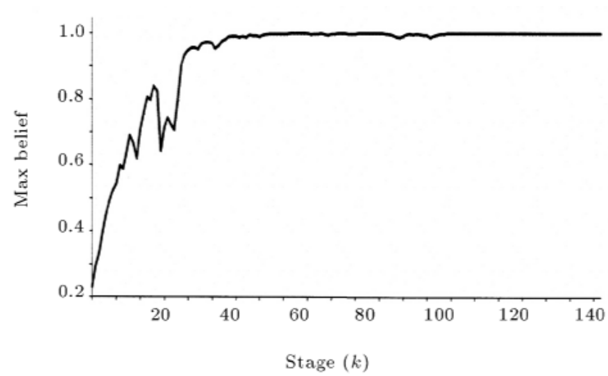

Figure 8. Converging process of f_4 (Example 8).

for determining N. Realizing the fact that the set of candidate functions, i.e., $S = f_1, f_2, \cdots f_m$, are known before, it will be possible to propose the following 3-step procedure for estimating N:

1. Generate random numbers for $\{f_1, f_2, \cdots f_m\}$ and compute their corresponding belief values $\{B_1(), B_2(), \cdots, B_m()\}$;
2. Compute $N^{\max} = \max\{N_i^{\max}$, for $i = 1, 2, \cdots m\}$ when $N_i^{\max}$ is the maximum number of observations needed for convergence of f_i;
3. Set $N = N^{\max}$.

Now, the working of the above procedure is illustrated in Example 9.

Example 9

Consider the following,

$$f_1 = \text{Gamma } (3,4), \qquad f_2 = \text{Gamma } (12,2),$$

$$f_3 = \text{Gamma } (16, \sqrt{3}), \qquad f_4 = \text{Gamma } (4, 2\sqrt{3}).$$

The functions are simulated and their associated belief values are calculated, as shown in Figure 9.

From the curves denoted as L1, L2, L3 and L4 in Figure 9 and, in accordance with Step 2 in the proposed procedure, one has $N_1^{\max} \approx 50$, $N_2^{\max} \approx 30$, $N_3^{\max} \approx 25$ and $N_4^{\max} \approx 50$. Hence, according to Step 3 in the proposed procedure, $N^{\max} = \max\{50, 30, 25, 50\} = 50$. Therefore, one can safely start the E-M algorithm by setting $N = 50$.

The above 3-step procedure for estimating N is only taken in a simulated environment and does not effect real experimentation, which may incur cost. It is considered that the proposed procedure should be used as a preprocessing step before application of the E-M algorithm.

It is also noticeable that the $N^{\max}$, as considered above, is an upper bound for $N_i^{\max}$, ensuring the convergence of all f_i's. In reality, however, the number of stages required for the unknown function may be shorter than $N^{\max}$, so that it is possible to update $N^{\max}$ adaptively. This means that, by collecting any new observation, $N_{\max}$ should be reestimated. This may lead, eventually, to $N_{\min}^{\max} = N_{i=gr}$ i.e., minimizing the total number of observations. In this case, however, the procedure can no longer be applied as a preprocessor but as an integral part of the E-M algorithm. The full development of such an adaptive algorithm is a subject for further research.

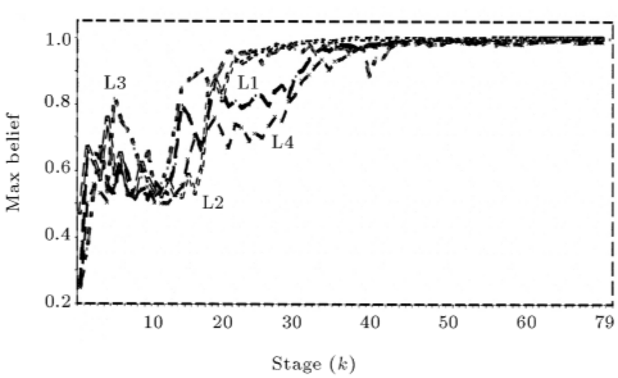

Figure 9. Simulation of converging process (Example 9).

Distribution Fit with Mixed Functions

Theoretically speaking, candidate functions in an E-M algorithm must be of a continuous type. This is naturally a limiting factor in application of the algorithm. In this section, an experiment is performed by implementing the method on a problem with both a continuous and a discrete nature, to see if it could work properly. Consider, $S = \{f_1, f_2, \cdots f_4\}$, f_1 = Exponential (1/8), f_2 = Poisson (10), f_2 = Poisson (8), f_2 = Poisson (6), $N = 12$, $V(N) = 0.95$ and $\alpha = 0.95$ where f_4 is the best fit function. Results are illustrated in Table 6.

As seen from the results illustrated in Table 6, it is clear that the algorithm still selects the best fit function correctly. However, theoretical difficulties may arise, which demand further investigation. (In a personal discussion with A. Eshragh Jahromi, he warranted the case that belief values, at any stage, may become equal, hence, stalling the process from further advancing. This is avoided in dealing with continuous functions.)

CONCLUSIONS AND FURTHER RESEARCH

Eshragh and Modarres [1-3] have developed a novel algorithm for a statistical estimation problem, called in this paper, the E-M algorithm. The algorithm uses a new sequential Bayesian method and a stochastic dynamical programming approach to determine when a process of obtaining observations can be stopped. Despite the originality and excellent mathematical analysis developed in the work, the presentation of the algorithm has been very difficult. The E-M algorithm has been presented by a new 3-phase method that illustrates the logical line of the algorithm and its implementation procedures. Finally, the results of our further developments have been resported.

Still, the algorithm can further be developed in some new directions. In order to predict the right candidate function at the present time, the only information being used is the value of x_k. However, it is quite plausible to introduce further information that can be deriven from a stream of x_k's, including mode, median, standard deviations and other distribution moments to accelerate the convergence of the algorithm

Table 6. Results for the mixed case.

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=9$	$K=10$	$K=11$
x_k	6	11	4	8	7	6	11	9
$B_1(x, O_{k\ 1})$	0.145	0.086	0.111	0.045	0.018	0.006	0.003	0.001
$B_2(x, O_{k\ 1})$	0.155	0.333	0.107	0.107	0.074	0.019	0.039	0.042
$B_3(x, O_{k\ 1})$	0.396	0.168	0.382	0.351	0.373	0.323	0.128	0.075
$B_4(x, O_{k\ 1})$	0.301	0.410	0.398	0.495	0.533	0.650	0.828	0.880
$sm,\ gr$	4,3	2,4	3,4	3,4	4,3	3,4	3,4	3,4
$B_{gr,sm}(gr, x_k, O_{k\ 1})$	0.568	0.551	0.510	0.584	0.588	0.668	0.865	0.920
$d^*(n)$	1	1	1	1	1	1	1	0.810
$V^*(n)$	0.659	0.679	0.700	0.722	0.744	0.841	0.867	0.893
Decision	-	-	-	-	-	-	-	f_4

by using fuzzy logic and neural networks [11,12]. This requires further investigations leading to development of a new parallel algorithm for distribution fitting problems [13].

ACKNOWLEDGMENT

The authors wish to thank the stimulating discussions with Ali Eshragh Jahromi and Dr. Modarres which helped to develope this paper to the present quality.

REFERENCES

1. Eshragh Jahromi, A. and Modarres Yazdi, M. "Decision on belief: A new approach to distribution fitting", *First National Conference on Industrial Engineering*, Sharif University of Technology, pp 233-283 (in Farsi) (2001).
2. Ranaeifar, F. "Critical Analysis of a new approach to distribution fitting: Decision on belief (D.O.B)", Undergraduate Thesis, Industrial Engineering Department, Alzahra University (see the report in http://staff.alzahra.ac.ir/saniee/undergraduatethesis) (2005).
3. Modarres Yazdi, M. and Eshragh Jahromi, A. "Decision on beliefs: A new approach to distribution fitting", Submitted to *Operations Research*.
4. Shafer, G., *A Mathematical Theory of Evidence*, Princeton (1976).
5. Dreyfus, S.E. and Law, A.M., *The Art and Theory of Dynamic Programming*, Academic Press, Inc. (1977).
6. Ross, S.M., *Introduction to Stochastic Dynamic Programming*, Academic Press, New York, USA (1983).
7. Ahn, J.H. and Kim, J.J. "Action-timing problem with sequential Bayesian belief revision process", *European Journal of Operational Research*, **105**, pp 118-129 (1998).
8. Conover, W.J., *Practical Nonparametric Statistics*, 2nd Ed., John Wiley & Sons (1980).
9. Industrial Lingo/P.C., Release 6.0, Lindo Systems, Inc (2000).
10. Monfared, M.A.S. and Yang, J.B. "Sensitivity analysis and parameter optimization", *Int. J. of Intelligent and Fuzzy Systems*, **15**(2), pp 89-104 (2004).
11. Monfared, M.A.S. and Steiner, S.T. "Fuzzy adaptive scheduling and control systems", *Int. J. Fuzzy Sets and Systems*, **115**, pp 231-246 (2000).
12. Monfared, M.A.S. and Yang, J.B. "Multi level intelligent scheduling and control system for an automated flow shop manufacturing environment", *Int. J. Production Research*, **43**(1), pp 147-168 (2005).
13. Monfared, M.A.S. and Khalifeh, F. "Following parallel paths: A new parallel processing algorithm to solve linear programs", *Amirkabir Journal of Engineering* (in Farsi) (in press).