A STOCHASTIC APPROXIMATION-ITERATIVE LEAST SQUARES ESTIMATION PROCEDURE

Ali H. Kashmar
Departement of Computer Science, Collage of Science, University of Baghdad. Baghdad-Iraq

Abstract
We consider the general nonlinear regression problem $Y(x) = g(\theta; x) + \varepsilon$. A survey of some classical methods and stochastic approximation procedures for estimating θ is first given. We solve the nonlinear regression problem by considering the optimal stochastic approximation procedure by [3],[4]. This leads us to introduce a new procedure , called "Stochastic Approximation Iterative Least Square Procedure" SA -ILS procedure. The new procedure is applied to a number of nonlinear regression models. We report on the results of a simulation investigation which indicate that the new procedure is highly efficient with respect to the number of observations required to obtain the parameter estimates for given regression problem.

التقريب العشوائي- بواسطة طريقة التقديرالأقل مربعات المكررة

علي حبيب كشمر
قسم علوم الحاسبات، كلية العلوم، جامعة بغداد. بغداد- العراق

الخلاصة
تم دراسة مشكلة الانحدار الغير خطي العام $Y(x) = g(\theta; x) + \varepsilon$. أعطينا أولا عرضا للمفاهيم الأساسية لبعض الطرق التقليدية وطرق التقريب العشوائي لتقدير θ. وبعدها أوجدنا حل المشكلة السابقة بواسطة طرق التقريب العشوائي المثلى المبتكرة بواسطة [3],[4] واقترحنا طريقة جديدة أطلقنا عليها "التقريب العشوائي – الأقل مربعات المكررة" وأعطيت الطريقة اسما مختصرا SA –ILS . لقد قمنا بتطبيق الطريقة المقترحة على عدد من نماذج الانحدار الغير خطي كما سجلنا نتائج دراسة المحاكاة التي تشير إلى أن الطريقة الجديدة ذات كفاءة عالية بالنسبة لعدد المشاهدات اللازمة لتقدير المعلمات في مشكلة الانحدار.

1-introduction

Let f be unknown function to the experimenter , and that for any level x we can observe a random variable Y(X) with expectation f(x). Let α be a given constant such that f(x)= α , has a unique root θ. The goal is to estimate θ, By choosing the appropriate measurement scale, we can without loss of generality assume that a = 0. [22] proposed a method for solution of this problem and a more general one, which is called the method of "Stochastic Approximation Procedure". Let the outcome of the measurement at x_n be $Y_i = f(x_i) + \varepsilon_i$, $i = 1,2, ...,$n where $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are independent random variables with $E(\varepsilon_n) = 0$ and $var(\varepsilon_n) = \sigma^2 < \infty$ for all n , and independent of x , and $f(x_n)$ can only be measured by an observer subject to random error ε_n whose magnitude cannot be neglected in view of the accuracy demanded of the solution f(x) =a. [22] assume $(x-\theta)f(x) > 0 \; for\ all\ x \neq \theta$, and under more conditions on f , proved that x_n converges in probability, and in mean square to θ .[17] considered the problem of estimating the value of θ sequentially such that f(x) achieves its minimum (or maximum). Let (a_n) and (c_n) be sequences of positive numbers such that $c_n \to 0$, $\sum_{n=1}^{\infty} a_n < \infty$, $\sum_{n=1}^{\infty} a_n c_n < \infty$, $\sum_{n=1}^{\infty} a_n^2 c_n^{-2} < \infty$ and assume $Y_n = \frac{Y_n^u - Y_n^1}{2}$, where Y_n^u and Y_n^1 are two observations. Let x_1 be an arbitrary initial value, then define the stochastic approximation procedure by: $x_{n+1} = x_n - a_n c_n^{-1} Y_n , n = 1,2,\ldots$ [17] proved, under certain conditions on f that x_n converges in probability to θ. [18] proved under weaker conditions than the conditions of [22],that x_n converges to θ almost surely. [15] proved, under weaker conditions than conditions of [8] on stochastic approximation procedure proposed by [22],that x_n converges with probability one to θ . [12] noticed some difficulties with the original stochastic approximation procedure $x_{n+1} = x_n - a_n Y_n , n = 1,2,\ldots$when $|f^*|$ was large near θ and small away from θ so that the correction $a_n Y_n$ tends to be too big when x_n was near θ , and otherwise it is too small [14]. He observed that Y_n in $x_{n+1} = x_n - a_n Y_n , n = 1,2,\ldots$ may change sign suddenly, so he proposed to take the absolute value of $c_n^{-1} Y_n$ multiplied by sign $c_n Y_n$, the new stochastic approximation procedure is as follows: choose x_1 as an arbitrary initial value of x , then define the estimating sequence x_n by: $x_{n+1} = x_n - a_n |c_n^{-1} Y_n| sign(c_n Y_n) , n = 1,2,\ldots$. He proved , under some additional conditions, that x_n converges to θ . Also [12] proposed another stochastic approximation procedure , by replacing $|c_n^{-1} Y_n| sign(c_n Y_n)$ by $c_n^{-1} \; sign(c_n Y_n)$, then the new stochastic approximation is $x_{n+1} = x_n - a_n c_n^{-1} \; sign(c_n Y_n) , n = 1,2,\ldots$. He proved, under some additional conditions that x_n converges to . Then question remaining, in general, what is the optimal transformation h which, when used in $x_{n+1} = x_n - a_n c_n^{-1} \; sign(c_n Y_n) , n = 1,2,\ldots$. will not destroy the convergence to θ and make the speed maximal. If the conditional distribution of $v_n = c_n Y_n - E_{(x_1 x_2,\ldots,x_n)} c_n Y_n$ (given$(x_1, x_2, \ldots, x_n)$) ; where v_n is a random variable ,distributed according to a distribution function F which is symmetric around 0 , and admits a density f , the optimal choice of h is $-c(f^*/f)$; where c is a constant ; if some additional mild requirements are satisfied. This result was obtained, independently, by [1] and [6]; the first considered both [22] and [17] situation, the second considered [22] situation.[3]proposed a general stochastic approximation procedure in the form $x_{n+1} = x_n - a_n c_n^{-1} A_n h_n , n = 1,2,\ldots,$ where h_n are designed r-vectors based on transforming the observations Y_n by a Boral measurable transformation h ,A_n are (q x r) measurable functions of $(x_1, h_1, \ldots, h_{n-1})$,$a_n$,c_n positive numbers , and x_1 is an arbitrary random vector in R^q . Under certain assumptions proposed by [2] , he established the almost sure convergence and the asymptotic normality results of the general stochastic approximation procedure $x_{n+1} = x_n - a_n c_n^{-1} A_n h_n , n = 1,2,\ldots$. Furthermore, the optimal choices of (a_n, c_n) are found by [3]. As for the optimal transformation, h_0 , is shown to be equal to $(I^{1/2}(f))^{-1}(-gradf(v)/f(v))$ a.e (F) with I(F) being the Fisher information matrix of f , provided that the error random vectors have a conditional (on$(x_1, x_2, \ldots x_n)$ distribution function F that admits a density function f whose gradient vector exists a.e. (F) with I(F) positive definite.

Consider the following general nonlinear regression model:

$$Y(x_1,x_2,.,x_r) = g(\theta_1,\theta_2,.,\theta_p;x_1,x_2,...x_r) + \varepsilon \quad (1.1)$$

Where $g: R^p \times R^r \to R$

with $= (x_1,x_2,....x_r)$; $\theta = (\theta_1,\theta_2,.....\theta_p)$, equation (1.1) can be written as:

$$Y(x) = g(\theta;x) + \varepsilon \quad (1.2)$$

Where ε is an unobservable centered random error , and its distribution may be dependent on x , but with $E(\varepsilon) = 0$ and $\sigma_\varepsilon^2(x) = \sigma^2$. Moreover, $Y(x)$ is an observable random variable at each x. Our interest will be in the class of models which contain a component linear in some parameters but nonlinear in the remaining parameters. The objective will be to estimate θ sequentially using a technique in which the optimal stochastic approximation method [4],is combined will the approach of eliminating linear parameters proposed by [18]. The sequential procedure is also compared with the fixed sample size procedure based fully on the [18] method. Let us first explain the procedure of [4], for the general model(1.2). consider the family of distribution functions $F^* = (F(.,\theta);\theta \in R^p)$ which is absolutely continuous with respect to a σ-finite measure μ on the Boreal σ-finite of R^r for some $r \geq 1$. The family F^* satisfies the following conditions:

(i): For $F(.,\theta) \in F^*, f(.,\theta)$ denotes its density function w .r .t μ;

(ii): The gradient of $f(.,\theta)$ w. r .t θ, ${}^{grad}{}_{\theta}f(.,\theta)$,exists for almost all θ (Lebesque) and $f(x,\theta)$ is measurable in both (x,θ), i.e., $f: R^p \times R^r \to R$ is measurable map;

(iii): The (p x p)-Fisher information matrix (θ), where

$$I_{ij}(\theta) = \int_{R^r} \frac{[\partial^2 f(x,\theta)/\partial\theta_i\partial\theta_j]}{f(x,\theta)}\mu(dx), i,j = 1,2,...,p$$

of $f(.,\theta)$ is positive definite for all $\theta \in \Theta \subset R^p$, so that its inverse $(I(\theta))^{-1}$ exists , where Θ is a subset of Euclidian P-space $R^p, p \geq 1$, and unknown vector of parameters θ is known to lie in Θ. Let Y be the vector random variable that has a density $f(.,\theta) \in F^*, \theta \in \Theta \subset R^p$,which is known except for the vector of parameters θ. Let $Y_1,Y_2,....$ be observations that will be drawn sequentially from Y that is defined on a probability space (Ω, F, p) with $PY^{-1} = P_\theta; \frac{dP_\theta}{du} = f(.,\theta) \in F^*$ .Define the following vector random variables:

$$Z_n = -(I(\theta^{(n)}))^{-1}[{}^{grad}{}_{\theta}f(x_n,\theta^{(n)})/f(x_n,\theta^{(n)})], n = 1,2,.$$

Thus Z_n can be considered as an observation on the family of vector random variables, $\{Z(\theta);\ \theta \in R^p\}$, defined by:

$$Z(\theta) = -(I(\theta^{(n)}))^{-1}[{}^{grad}{}_{\theta}f(x_n,\theta\ )/f(x,\theta)]$$

Moreover define the following Boral measurable regression function:

$$M(\theta) = E(Z(\theta)) = -(I(\theta))^{-1}\int_{R^r}[{}^{grad}{}_{\theta}f(x,\theta)/f(x,\theta)]f(x,\theta)\mu(dx)$$

which exists for all $\theta \in \Theta \subset R^p$,.Now to achieve our objective , i.e., to estimate θ sequentially , we can then use the following optimal stochastic approximation procedure [3] [4], choose $\theta^{(1)}$ as an arbitrary initial estimate of θ , then define the estimating sequence $\theta^{(n)}$ by :

$$\theta^{(n+1)} = \theta^{(n)} - a_n h_o(Y_n)\ ,\ n = 1,2,.... (1.3)$$

Where

$$h_o(Y_n) = [-(I_{-n}(\theta))^{-1}[{}^{grad}{}_{\theta}f(Y_n,\theta)/f(Y_n,\theta)]]_{\theta=\theta^{(n)}} \quad ,n = 1,2,.$$

and a_n is a sequence satisfying $\sum_{n=1}^{\infty} a_n = \infty,\ \ \sum_{n=1}^{\infty} a_n^2 < \infty$

[18] of the estimation may be applied when the nonlinear regression model(1.2) has the special form $Y(x) = \sum_{j=1}^{q}\theta_{i,j}g_j(\theta_{(2)};x) + \varepsilon$ (1.4)

Where $\theta_{1,1},\theta_{1,2},....,\theta_{1,q}$ enter linearly into the model (1.4), $\theta_{(2)}$ represents the vector of nonlinear parameters in (1.4),and, the $g_j(\theta_{(2)};x)$ are functions only of the nonlinear parameters and the predictor variables, i.e., $g_j: R^{p-q} \times R^r \to R$, $j = 1,2,...,q$. Their method is used for the fixed sample size case when observations $Y_1,Y_2,....Y_n$ are available. Using their procedure, we take $\theta_{(2)}$ as an initial value of $\theta_{(2)}$, and then determine the companion set of "best" values for $\theta_{(1)}$ by the ordinary least squares procedure. Let

$\theta_{(1)}(\theta_{(2)}) = (\theta_{1,1}(\theta_{(2)}), \theta_{1,2}(\theta_{(2)}), \ldots\ldots \theta_{1,q}(\theta_{(2)}))$ represent the vector of least squares estimates of the $\theta_{1,j}'s$ associated with a given set of $\theta_{(2)}$'s ; namely , $\theta_{(2)} = (\theta_{2,1}, \ldots \theta_{2,p-q})$ .Let Y denote the (nx1) column vector of observed response values associated with the n observed value of the predictor vector, x_i , i=1,2,…,n. Let $G_{\theta_{(2)}}$ denote the (nxq) matrix with elements $g_j(\theta_{(2)}; x_i), i = 1,2,..n, j = 1,2,\ldots q$ it then follows that the vector $\theta_1(\theta_{(2)})$; provided that $(G_{\theta_{(2)}} G_{\theta_{(2)}})^{-1}$ exists; is given by: $\theta_1(\theta_{(2)}) = (G_{\theta_{(2)}} G_{\theta_{(2)}})^{-1} G_{\theta_{(2)}} Y$ .The reduced "model" associated with (1.4) is then given by:

$$Y(x) = \sum_{j=1}^{q} \theta_{i,j}(\theta_{(2)}) g_j(\theta_{(2)}; x) + \varepsilon^* \quad (1.5)$$

Since the $\theta_{i,j}(\theta_{(2)})$ are strictly functions of $\theta_{(2)}$'s , the model in (1.5) is a nonlinear regression model with only (p-q) parameters rather than the P parameters in the original model. [18] proposed to estimate the remaining unknown nonlinear parameters by using an iterative method like the linearization method , steepest descent method or other known methods in the literature. This procedure of [18] estimate the nonlinear parameters in a nonsequentially fashion, that is the whole data must be used to find values of the estimators. If the data is drawn sequentially, then these procedure will not be suitable to use. In addition the procedures introduced by [18] are not "optimal" in some sense. However the stochastic approximation procedures have been shown to be 'optimal" [4] in the sense that the estimating sequence $(\theta^{(n)})$ is a consistent and asymptotically efficient estimator of θ , such that, the variance of the asymptotic distribution of $n^{1/2}(\theta^{(n)} - \theta)$ achieves the Gramer-Rao lower bound for the variance of unbiased estimator of θ . The above results show that it is worthwhile to consider the use of stochastic approximation procedures to estimate sequentially the nonlinear parameters in (1.5), instead of using any iterative classical method.

2-Illustration of the Lawton and Syivestre Procedure

We will illustrate the fixed sample size procedure of Lawton and Syivestre [18] using the following example given by these authors. Let

$$g(\theta_1, \theta_2; x) = \theta_1 e^{\theta_2 x} \quad (2.1)$$

i.e., $Y(x) = \theta_1 e^{\theta_2 x} + \varepsilon$ (2.2)

where θ_1 and θ_2 are two unknown parameters to be estimated, ε is an unobservable random error and Y(x) is a response variable at the level x . θ_1 appears linearly in the model (2.2). We seek the least squares estimators θ_1, θ_2 which minimize

$$Q(\theta_1, \theta_2) = \sum_{i=1}^{n} (y_1 \quad \theta_1 e^{\theta_2 x_i})^2 \quad (2.3)$$

For θ_2 fixed at any value , a partial minimum for $Q(\theta_1, \theta_2)$ is obtained by setting

$$\frac{\partial Q(\theta_1, \theta_2)}{\partial \theta_1} = \frac{\partial}{\partial \theta_1} \sum_{i=1}^{n} (Y_i - \theta_1 e^{\theta_2 x_i})^2 = 0 \quad (2.4)$$

Denoting this best value of θ_1 given θ_2 by $\theta_1(\theta_2)$,(2.4) yields

$$\theta_1(\theta_2) = \frac{\sum_{i=1}^{n} (Y_i e^{\theta_2 x_i})}{\sum_{i=1}^{n} e^{2\theta_2 x_i}} \quad (2.5)$$

Now substitute (2.3) into (2.5). The linear parameter θ_1 is automatically replaced by its best companion value $\theta_1(\theta_2)$ which is a function of θ_2 alone. One then obtains the reduced "model", given by:

$$Y(x) = \theta_1(\theta_2) e^{\theta_2 x} + \varepsilon^* \quad (2.6)$$

The parameter θ_2 will be estimated iteratively by using any of the iterative method (Linearization, Steepest Descent, Marquardt's Compromise)[11,21,20,9,18] gave two benefits for this parameter reduction procedure

(i): Convergence seem to be faster, than other procedures that do not employ the reduction technique, (but no proof is furnished), and more stable because of reduction dimensionality of the parameter space

(ii): One has to supply only an initial guess for the reduced number of parameters.A number of important applications of the model being considered were described by [18],[19], in connection with experiments in spectrophotometers.

3-Stochastic Approximation-Iterative Least Squares Procedure (SA – ILS)

We will describe a new sequential procedure for estimating the parameters in the model given by (1.4), which combines the Stochastic Approximation technique with the Iterative Least Squares technique. For abbreviation this will be referred to , as the (SA-ILS) procedure. Clearly , the reduced "model" in (2.6) is a nonlinear model with a single parameter θ_2 . In order to estimate θ_2 sequentially by using optimal stochastic approximation procedure. We shall consider certain probability model for ε , by using the reduced "model" in (2.6), and then find the probability density function for Y , $f(y;\theta_2)$, by transformation. Thus , by using optimal stochastic approximation procedure of the form (1.3), in order o estimate θ_2 sequentially , choose $\theta_2^{(1)}$ as an arbitrary initial estimate of θ_2 , then define the estimating sequence $\theta_2^{(n)}$ by :

$$\theta_2^{(n+1)} = \theta_2^{(n)} + a_n[I_n(\theta_2^{(n)})]^{-1}[\frac{df(Y_n;\theta_2)/d\theta_2}{f(Y_n;\theta_2)}]_{\theta_2=\theta_2^{(n)}}, n = 1,2,\ldots \quad (3.1)$$

Where $a_n = 1/n$, since the optimal value of a that minimizes the variance of the asymptotic distribution of $n^{\frac{1}{2}}(\theta_2^{(n)} - \theta_2)$,[4] is given by 1 , noting that for notational convenience we set $f_n\left(Y_n;\theta_1^{(n)},\theta_2\right) = f(Y_n;\theta_2)$. Then (3.1) will become:

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{1}{n}[I_n(\theta_2^{(n)})]^{-1}[\frac{df(Y_n;\theta_2)/d\theta_2}{f(Y_n;\theta_2)}]_{\theta_2=\theta_2^{(n)}}, n = 1,2,\ldots \quad (3.2)$$

The main idea of SA-ILS procedure is to estimate the parameters which enter the model linearly, by using an iterative form of least squares estimators, sequentially, and then use a proper optimal stochastic approximation procedure to sequentially estimate nonlinear parameters. Therefore , we will use iterative least squares procedure in order to estimate $\theta_1(\theta_2)$ sequenatially. To construct the general formal for the sequence $(\theta_1(\theta_2))$, we have: $\theta_1(\theta_2) = \frac{\sum_{i=1}^{n}(Y_i e^{\theta_2 x_i})}{\sum_{i=1}^{n} e^{2\theta_2 x_i}}$,. For n=1

Given an initial guess $\theta_2^{(1)}$, we then have $\theta_1^{(1)}\left(\theta_2^{(1)}\right) = \frac{Y_1 e^{\theta_2^{(1)} x_1}}{e^{2\theta_2^{(1)} x_1}} = \frac{Y_1}{e^{\theta_2^{(1)} x_1}}$, since $\theta_1^{(1)}\left(\theta_2^{(1)}\right) = \theta_1^{(1)}$, then $\theta_1^{(1)} = \frac{Y_1}{e^{\theta_2^{(1)} x_1}}$,. For n=2: Substitute $\theta_1^{(1)}$ in (3.2) , we will get $\theta_2^{(2)}$, and then $\theta_1^{(1)}\left(\theta_2^{(1)}\right) = \frac{\sum_{i=1}^{2} Y_1 e^{\theta_2^{(2)} x_1}}{\sum_{i=1}^{2} e^{2\theta_2^{(2)} x_1}}$, which may write as

$$\theta_1^{(2)}\left(\theta_2^{(2)}\right) = \theta_1^{(1)}\left(\theta_2^{(1)}\right) + \frac{1}{\sum_{i=1}^{2} e^{2\theta_2^{(2)} x_1}}[\sum_{i=1}^{2} Y_1 e^{\theta_2^{(2)} x_1} - \theta_1^{(1)}\left(\theta_2^{(1)}\right)\sum_{i=1}^{2} e^{2\theta_2^{(2)} x_1}]$$

i.e.,

$$\theta_1^{(2)} = \theta_1^{(1)} + \frac{1}{\sum_{i=1}^{2} e^{2\theta_2^{(2)} x_1}}[\sum_{i=1}^{2} Y_1 e^{\theta_2^{(2)} x_1} - \theta_1^{(1)}\sum_{i=1}^{2} e^{2\theta_2^{(2)} x_1}]$$

in general, at state n , by substituting $\theta_1^{(n-1)}$ in (3.2) ,we get $\theta_2^{(n)}$, and then

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n} e^{2\theta_2^{(n)} x_1}}\left[\sum_{i=1}^{n} Y_1 e^{\theta_2^{(n)} x_1} - \theta_1^{(n-1)}\sum_{i=1}^{n} e^{2\theta_2^{(2)} x_1}\right], n = 1,2,\ldots$$

(3.3)

Where $\theta_1^{(1)} = \frac{Y_1}{e^{\theta_2^{(1)} x_1}}$, will be an initial estimate of $\theta_1^{(n)}$, given an initial value $\theta_2^{(1)}$ of θ_2 . Therefore the SA-ILS procedure is given by the following two consecutive procedures:

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{1}{n}[I_n(\theta_2^{(n)})]^{-1}[\frac{df(Y_n;\theta_2)/d\theta_2}{f(Y_n;\theta_2)}]_{\theta_2=\theta_2^{(n)}}, n = 1,2,\ldots$$

and

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n} e^{2\theta_2^{(n)} x_1}}\left[\sum_{i=1}^{n} Y_1 e^{\theta_2^{(n)} x_1} - \theta_1^{(n-1)}\sum_{i=1}^{n} e^{2\theta_2^{(2)} x_1}\right], n = 1,2,\ldots$$

where $\theta_2^{(1)}$ is an arbitrary initial value for the sequence $\theta_2^{(n)}$ and $\theta_1^{(1)} = \frac{Y_1}{e^{\theta_2^{(1)} x_1}}$ is the initial estimate of ($\theta_1^{(n)}$) based on $\theta_2^{(1)}$.To explain the SA-ILS computational procedure , we will demonstrate how to compute the first three estimates:

<u>Step 1</u>: Let $\theta_2^{(1)}$ be an arbitrary initial estimate of θ_2 before any data are collected.

Step 2 : For $\theta_2 = \theta_2^{(1)}$ and data (Y_1, X_1), the value of θ_1 which minimizes $(Y_1 - \theta_1 e^{\theta_2^{(1)} x_1})^2$ is obtained as $\theta_1^{(1)} = \frac{Y_1}{e^{\theta_2^{(1)} x_1}} / e^{2\theta_2^{(1)} x_1} = Y_1 / e^{\theta_2^{(1)} x_1}$

Step 3 : Treating $\theta_1^{(1)}$ as if it was the known true value of θ_1, the second estimate of θ_2 is obtained from

$$\theta_2^{(2)} - \theta_2^{(1)} + [I_1(\theta_2^{(1)})]^{-1} [\frac{df(Y_1;\theta_2)/d\theta_2}{f(Y_1;\theta_2)}]_{\theta_2 = \theta_2^{(1)}}$$

Step 4 : For $\theta_2 = \theta_2^{(2)}$ and data $(Y_1, X_1), (Y_2, X_2)$, the value of θ_1 which minimize $\sum_{n=1}^{2} (Y_n - \theta_1 e^{\theta_2^{(2)} x_n})^2$ is obtained as:

$$\theta_1^{(2)} = \theta_{(1)}^{(n)}$$

$$\theta_1^{(1)} + \frac{1}{\sum_{n=1}^{2} e^{2\theta_2^{(2)} x_n}} \left[\sum_{n=1}^{2} Y_n e^{\theta_2^{(2)} x_n} - \theta_1^{(1)} \sum_{n=1}^{2} e^{2\theta_2^{(2)} x_n} \right]$$

Step 5 : Treating $\theta_1^{(2)}$ as if it was the new known true value of θ_1, the third estimate of θ_2 is obtained from

$$\theta_2^{(3)} = \theta_2^{(2)} + \frac{1}{2} [I_2(\theta_2^{(2)})]^{-1} [\frac{df(Y_2;\theta_2)/d\theta_2}{f(Y_2;\theta_2)}]_{\theta_2 = \theta_2^{(2)}}$$

These above steps are repeated until convergence occurs with $|\theta_2^{(n+1)} - \theta_2^{(n)}| < \delta$, where δ is a small specified positive number. In the vector case, to estimate $\theta_{(1)}$ sequentially by applying an iterative least squares procedure to the model (1.4), then given $\theta_{(2)}^{(1)}$ as an arbitrary initial value to initiate the sequence $(\theta_{(2)}^{(n)})$, one can determine the companion set of "best" values for the q linear parameters by linear regression. Let

$\theta_1^{(n)}(\theta_{(2)}) = (\theta_{1,1}^{(n)}(\theta_{(2)}), \dots \dots \theta_{1,q}^{(n)}(\theta_{(2)}))'$

represent the vector of iterative least squares estimates. Let Y_1 be the first observation, Y_2 be the vector of the first two observations, and so on , Y_n denoting the vector of the first n observed response values which have associated observed values of the predictor vector , X_i , $i = 1,2, \dots, n$. Let $G_{\theta_{(2)}^{(n)}}$ denote the (n x q) matrix with elements $g_j\left(\theta_{(2)}^{(n)}; X_i\right)$, $i = 1,2,\dots,n$, $j = 1,2, \dots, q$

. It then follows that the sequence $(\theta_{(1)}^{(n)}(\theta_{(2)}^{(n)}))$, provided that $(G_{\theta_{(2)}^{(n)}} G_{\theta_{(2)}^{(n)}})^{-1}$ exists, is given by: $\theta_{(1)}^{(n)}(\theta_{(2)}^{(n)}) = \theta_{(1)}^{(n-1)}(\theta_{(2)}^{(n-1)}) + [(G_{\theta_{(2)}^{(n)}} G_{\theta_{(2)}^{(n)}})^{-1}(G_{\theta_{(2)}^{(n)}} Y_n - (\theta_{(1)}^{(n-1)}(\theta_{(2)}^{(n-1)})]$,

$n = 1,2,\dots$ and $\theta_{(1)}^{(n)}(\theta_{(2)}^{(n)}) = \theta_{(1)}^{(n)}$, i.e.,

$$= \theta_{(1)}^{(n-1)} + [(G_{\theta_{(2)}^{(n)}} G_{\theta_{(2)}^{(n)}})^{-1}(G_{\theta_{(2)}^{(n)}} Y_n - \theta_{(1)}^{(n-1)}], n = 1,2,\dots \quad (3.4)$$

Where $\theta_{(2)}^{(1)}$ is an arbitrary initial value for the sequence $(\theta_{(2)}^{(n)})$, and $\theta_{(1)}^{(1)} = (G_{\theta_{(2)}^{(1)}} G_{\theta_{(2)}^{(1)}})^{-1}(G_{\theta_{(2)}^{(1)}} Y_1$ is an initial value for the sequence $(\theta_{(1)}^{(n)})$

The reduced model associated with (1.4) is then given by ;

$$Y(x) = \sum_{j=1}^{q} \theta_{i,j}(\theta_{(2)}) g_j(\theta_{(2)}; x) + \varepsilon^* \quad (3.5)$$

Since the $\theta_{i,j}(\theta_{(2)})$ are strictly functions of the $\theta_{(2)}$' s ; the model (3.5) is a nonlinear regression model with only (p-q) parameters, and we will estimate them by using optimal stochastic approximation procedure of the form (1.3).

4- Examples of the use of the SA-ILS procedure under different error distributions

We shall consider the following nonlinear regression functions;

Example 1 : $Y(x) = \theta_1 e^{\theta_2 x} + \varepsilon = g(\theta_1, \theta_2; x) + \varepsilon$ [18]

Example 2: $Y(x) = \theta_1 \sin\theta_2 x + \varepsilon = g(\theta_1, \theta_2; x) + \varepsilon$ [5]

Example 3 : $Y^*(x_1,x_2) = \exp\left\{-\theta_1 x_1 \exp\left[-\theta_2\left(\frac{1}{x_2} - \frac{1}{620}\right)\right]\right\} \times \varepsilon^* = g(\theta_1,\theta_2,x_1,x_2) \times \varepsilon^*$ [11]

Also , we shall consider the following three probability models for ε ;

(4.1) $\varepsilon \sim N(0,1)$

(4.2) $\varepsilon \sim$ Double exponential distribution

(4.3) $\varepsilon \sim$ T-distribution with r degree of freedom ≥ 1 , which includes the Cauchy distribution (r=1).First of all we will explain in an analytical form the steps of the procedure for each example separately. Then we describe the computer simulation investigation performed using VAX 11/785 Computer System VMS 4.2. We will apply three probability models for , on each example separately.

Example 1 under (4.1)

The probability density function of ε is given by: $f(\varepsilon) = \frac{1}{\sqrt{2\pi}} e^{-\varepsilon^2/2}$; $-\infty < \varepsilon < \infty$

Thus Y is also distributed as $N(g(\theta_1,\theta_2;x);1)$, that is , $Y \sim N(\theta_1 e^{\theta_2 x};1)$. Treating θ_1 as known initially , we have:

$$f(Y,\theta_1,\theta_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(Y - \theta_1 e^{\theta_2 x})^2\right) ; -\infty < Y < \infty$$

.Differentiating $\ln \ f(Y,\theta_1,\theta_2)$ w.r.t, θ_2 then we get:

$$\ln \ f(Y,\theta_1,\theta_2) = -\frac{1}{2}(Y - \theta_1 e^{\theta_2 x})^2) - \frac{1}{2}\ln(2\pi)$$

$$\frac{d \ln f(Y,\theta_1,\theta_2)}{d\theta_2} = X\theta_1 e^{\theta_2 x}(Y - \theta_1 e^{\theta_2 x}), -\infty < Y < \infty$$

The Fisher information, (θ_2) , is

$$I(\theta_2) = E_Y\left(\frac{d \ln f(Y,\theta_1,\theta_2)}{d\theta_2}\right)^2 = E_Y((X\theta_1 e^{\theta_2 x})^2 (Y - \theta_1 e^{\theta_2 x})^2) = \left((X\theta_1 e^{\theta_2 x})^2 E_Y(Y - \theta_1 e^{\theta_2 x})^2\right) = (X\theta_1 e^{\theta_2 x})^2 \qquad (4.1)$$

In view of (4.1) , we get $(I(\theta_2))^{-1} = \left(\frac{d \ln f(Y,\theta_1,\theta_2)}{d\theta_2}\right) = \frac{(Y - \theta_1 e^{\theta_2 X})}{\theta_1 X e^{\theta_2 x}}$; $X \neq 0$

From section (1.4) the optimal transformation for the stochastic approximation procedure is

$$h_0(Y_n) = -(I_n(\theta_2))^{-1}\left(\frac{d \ln f(Y,\theta_1,\theta_2)}{d\theta_2}\right); n = 1,2,\ldots$$

$$= \frac{-(Y_n - \theta_1(\theta_2) e^{\theta_2 X_n})}{(\theta_1(\theta_2) X_n e^{\theta_2 X_n})} \ ; \ X_n \neq 0, n = 1,2,\ldots.$$

.The optimal value of a that minimizes the variance of asymptotic distribution of $n^{1/2}(\theta_2^{(n)} - \theta_2)$ is give by 1 [4] ,then $a_n = 1/n$

Now choose $\theta_2^{(1)}$ as an arbitrary initial estimate of θ_2 , then define the estimating sequence $(\theta_2^{(n)})$ by:

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{1}{n} \frac{(Y_n - \theta_1^{(n)}(\theta_2^{(n)}) e^{\theta_2^{(n)} X_n})}{(\theta_1^{(n)}(\theta_2^{(n)}) X_n e^{\theta_2^{(n)} X_n})}, X_n \neq 0, n = 1,2,\ldots.$$

And the estimating sequence of $(\theta_1^{(n-1)})$ is given by:

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n} e^{2\theta_2^{(n)} X_i}}\left[\sum_{i=1}^{n} Y_i e^{\theta_2^{(n)} X_i} - \theta_1^{(n-1)} \sum_{i=1}^{n} e^{2\theta_2^{(n)} X_i}\right], n = 1,2,\ldots$$

Where $\theta_1^{(1)} = Y_1 / e^{\theta_2^{(1)} X_1}$

__Example 1 under (4.2)__

The probability density function of ε is given by : $f(\varepsilon) - \frac{1}{2} e^{-|\varepsilon|}, -\infty < \varepsilon < \infty$

Thus the density of Y is $f(y;\theta_1,\theta_2) = \frac{1}{2} e^{-|Y - \theta_1 e^{\theta_2 x}|}, -\infty < Y - \theta_1 e^{\theta_2 x} < \infty$

Treating θ_1 as known initially and differentiating $f(y;\theta_1,\theta_2)$ w .r. t θ_1 then we get:

$$\frac{df(Y;\theta_1,\theta_2)}{d\theta_2} = \begin{cases} \frac{1}{2}\theta_1 X e^{\theta_2 x} \exp\left(-(Y - \theta_1 e^{\theta_2 x})\right) ; 0 < Y - \theta_1 e^{\theta_2 x} < \infty \\ -\frac{1}{2}\theta_1 X e^{\theta_2 x} \exp(Y - \theta_1 e^{\theta_2 x}) ; -\infty < Y - \theta_1 e^{\theta_2 x} \leq 0 \end{cases}$$

And hence,

$$\frac{df(Y;\theta_1,\theta_2)/d\theta_2}{f(Y;\theta_1,\theta_2)} = \begin{cases} \theta_1 X e^{\theta_2 x} ; \ 0 < Y - \theta_1 e^{\theta_2 x} < \infty \\ -\theta_1 X e^{\theta_2 x} ; -\infty < Y - \theta_1 e^{\theta_2 x} \leq 0 \end{cases}$$

$$= \theta_1 X e^{\theta_2 x} \operatorname{sign}(Y - \theta_1 e^{\theta_2 x})$$

The Fisher information, $I(\theta_2)$ is :

$I(\theta_2) = E_Y(\frac{\frac{df(Y;\theta_1,\theta_2)}{d\theta_2}}{f(Y;\theta_1,\theta_2)})^2 =$

$E_Y(\theta_1 X e^{\theta_2 x}\,\mathrm{sign}(Y-\theta_1 e^{\theta_2 x}))^2$

$= (\theta_1 X e^{\theta_2 x})^2\, E_Y(sign(Y-\theta_1 e^{\theta_2 x}))^2$

$= (\theta_1 X e^{\theta_2 x})^2$ (4.2)

In view of (4.2), we get $(I(\theta_2))^{-1}(\frac{\frac{df(Y;\theta_1,\theta_2)}{d\theta_2}}{f(Y;\theta_1,\theta_2)}) = \frac{sign(Y-\theta_1 e^{\theta_2 x})}{\theta_1 X e^{\theta_2 x}}; X \neq 0$

Form section (1.4) the optimal transformation for the stochastic approximation procedure is $h_n(Y_n) = \frac{-sign(Y_n-\theta_1(\theta_2)e^{\theta_2 x_n})}{(\theta_1(\theta_2)x_n e^{\theta_2 x_n})}$; $n = 1,2,\ldots$; $x_n \neq 0$

Let $\theta_2^{(1)}$ be an arbitrary initial estimate of θ_2, then define the estimating sequence $(\theta_2^{(n)})$ by :

$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{sign\left(Y_n-\theta_1^{(n)}(\theta_2^{(n)})e^{\theta_2^{(n)}x_n}\right)}{n(\theta_1^{(n)}(\theta_2^{(n)})X_n e^{\theta_2^{(n)}x_n})}$; $n = 1,2,\ldots$; $x_n \neq 0$

And , the estimating sequence $(\theta_1^{(n+1)})$ is given by:

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n} e^{2\theta_2^{(n)}x_i}}\left[\sum_{i=1}^{n} Y_i e^{\theta_2^{(n)}x_i} - \theta_1^{(n-1)}\sum_{i=1}^{n} e^{2\theta_2^{(n)}x_i}\right], n = 1,2,\ldots$$

Where $\theta_1^{(1)} - Y_1/e^{\theta_2^{(1)}X_1}$

Example 1 under (4.3)

The probability density function of ε is given by

$f(\varepsilon) - \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{\pi r}}(1+\varepsilon^2/r)^{\frac{-(1+r)}{2}}, \quad -\infty < u < \infty$

Thus the density of Y is

$f(Y;\theta_1,\theta_2) = \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{\pi r}}(1+(Y-\theta_1 e^{\theta_2 x}))^2/r)^{\frac{-(1+r)}{2}}, \quad -\infty < Y < \infty$

Treating θ_1 as known initially , and for simplification letting $C = \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{\pi r}}$ we have :

$f(Y;\theta_1,\theta_2) = C(1+Y-\theta_1 e^{\theta_2 x})^2/r)^{\frac{-(1+r)}{2}}, \quad -\infty < Y < \infty$

Weobtain

$\frac{df(Y;\theta_1,\theta_2)}{d\theta_2} = C\left(\frac{1+r}{r}\right)\left(Y-\theta_1 e^{\theta_2 x}\right)\left(Xe^{\theta_2 x}\right)(1+Y-\theta_1 e^{\theta_2 x})^2/r)^{\frac{-(1+r)}{2}-1}$

, and,

$I(\theta_2) = E_Y(\frac{\frac{df(Y;\theta_1,\theta_2)}{d\theta_2}}{f(Y;\theta_1,\theta_2)})^2$

$= E_Y((\frac{1+r}{r})(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^{-1}(Y-\theta_1 e^{\theta_2 x})\theta_1 X e^{\theta_2 x})^2$

$= (\frac{\theta_1 x e^{\theta_2 x}(1+r)}{r})^2 E_Y(\frac{(Y-\theta_1 e^{\theta_2 x})^2}{(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^2})$

$= (\frac{\theta_1 x e^{\theta_2 x}(1+r)}{r})^2$

$$\int_{-\infty}^{\infty}\frac{(Y-\theta_1 e^{\theta_2 x})^2\Gamma\left(\frac{r+1}{2}\right)(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^{\frac{-(1+r)}{2}}}{(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^2\Gamma\left(\frac{r}{2}\right)\sqrt{\pi r}}dy$$

Let $= \frac{2(\theta_1 X e^{\theta_2 x}(r+1))^2\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})r^2\sqrt{\pi r}}$, since the integrand is symmetric about $\theta_1 e^{\theta_2 x}$,

then, $I(\theta_2) = R\int_{\theta_1 e^{\theta_2 x}}^{\infty}\frac{(Y-\theta_1 e^{\theta_2 x})^2}{(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^{(5+r)/2}}dy$ .Using integration by parts, we have

$V = (Y-\theta_1 e^{\theta_2 x}),\ du = (Y-\theta_1 e^{\theta_2 x})(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^{-(5+r)/2}dy$

$dv - dy \quad , \quad u - \frac{-r}{r+3}(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^{-(r+3)/2}$

thus,

$I(\theta_2) = R[\frac{-r(Y-\theta_1 e^{\theta_2 x})}{(r+3)(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^{(r+3)/2}}|_{(\theta_1 e^{\theta_2 x})}^{\infty} + r/r+3$

$\times\int_{(\theta_1 e^{\theta_2 x})}^{\infty}(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^{-(r+3)/2}dy$

$= \frac{R\,r}{r+3}\int_{(\theta_1 e^{\theta_2 x})}^{\infty}(1+(Y-\theta_1 e^{\theta_2 x})^2/r)^{-(r+3)/2}dy$

Putting $\frac{Y-\theta_1 e^{\theta_2 x}}{\sqrt{r}} = \tan\theta$ for which $dy = \sqrt{r}sec^2\theta d\theta$, we obtain

$I(\theta_2) = \frac{R\,r}{r+3}\int_0^{\pi/2}(1+(tan2\theta)^{-(r+3)2}\sqrt{r}sec^2\theta d\theta$

$= \frac{R\,r^{3/2}}{r+3}\int_0^{\pi/2}(cos)^{(r+1)}\theta d\theta$

It is straight forward to show that an alternative representation for $I(\theta_2)$ is

$$I(\theta_2) = \frac{R\, r^{5/2}}{(r+3)(r+1)}\int_0^{\pi/2}(cos)^{(r-1)}\theta d\theta\,, \quad (4.3)$$

Which has the advantage of a smaller power in the integrand. Using (4.3),we get

$$(I(\theta_2))^{-1}\frac{\frac{df(Y;\theta_1,\theta_2)}{d\theta_2}}{f(Y;\theta_1,\theta_2)} = \frac{(Y-\theta_1 e^{\theta_2 x})(r+3)\sqrt{n}\Gamma\left(\frac{r}{2}\right)}{2(\theta_1 X e^{\theta_2 x})(r+(Y-\theta_1 e^{\theta_2 x})^2)\Gamma\left(\frac{r+1}{2}\right)} \times [\int_0^{\frac{\pi}{2}}(cos)^{(r-1)}\theta d\theta]^{-1}, x \neq 0$$

From section (1.4) the optimal transformation for the stochastic approximation procedure is

$$h_0(Y_n) = \frac{-(Y_n-\theta_1(\theta_2)e^{\theta_2 x_n})(r+3)\sqrt{n}\Gamma\left(\frac{r}{2}\right)[\int_0^{\frac{\pi}{2}}(cos)^{(r-1)}\theta d\theta]^{-1}}{2(\theta_1(\theta_2)x_n e^{\theta_2 x_n})(r+(Y_n-\theta_1(\theta_2)e^{\theta_2 x_n})^2)\Gamma\left(\frac{r+1}{2}\right)}, x_n \neq 0, n = 1,2,\ldots, r \geq 1$$

Let $\theta_2^{(1)}$ be an arbitrary initial estimate of θ_2 , then define the estimating sequence $(\theta_2^{(n)})$ by :

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{\left(Y_n - \theta_1^{(n)}(\theta_2^{(n)})e^{\theta_2^{(n)}x_n}\right)(r+3)\sqrt{n}\Gamma\left(\frac{r}{2}\right)[\int_0^{\frac{\pi}{2}}(cos)^{(r-1)}\theta d\theta]^{-1}}{2n(\theta_1^{(n)}(\theta_2^{(n)})x_n e^{\theta_2^{(n)}x_n})(r+(Y_n-\theta_1^{(n)}(\theta_2^{(n)})e^{\theta_2^{(n)}x_n})^2)\Gamma\left(\frac{r+1}{2}\right)} \quad (4.4)$$

$x_n \neq 0, n = 1,2,\ldots, r \geq 1$, and , the estimating sequence $(\theta_1^{(n-1)})$ is given by :

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^n e^{2\theta_2^{(n)}X_i}}[\sum_{i=1}^n Y_i e^{\theta_2^{(n)}X_i} - \theta_1^{(n-1)}\sum_{i=1}^n e^{2\theta_2^{(n)}X_i}], n = 1,2,\ldots \quad (4.5)$$

Where $\theta_1^{(1)} = Y_1/e^{\theta_2^{(1)}X_1}$.Taking different cases of degrees of freedom r , the integral in (4.4) can be shown to be $\pi/2$, 1 , 2/3, 384/945 for r = 1, 2, 4, 10 respectively. The estimating sequence $(\theta_2^{(n)})$ in each case has the form

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{\lambda_r\left(Y_n - \theta_1^{(n)}(\theta_2^{(n)})e^{\theta_2^{(n)}x_n}\right)}{n(\theta_1^{(n)}(\theta_2^{(n)})x_n e^{\theta_2^{(n)}x_n})(r+(Y_n-\theta_1^{(n)}(\theta_2^{(n)})e^{\theta_2^{(n)}x_n})^2)}$$

, n = 1, 2, … ; $x_n \neq 0$,

Where $\lambda_r = 4,5,7,13$ for r = 1, 2, 4, 10 respectively. In each case the estimating sequence $(\theta_2^{(n-1)})$ is given by (4.5). The above coefficient values for λ_r suggest that he general form is $\lambda_r = r+3$, but we have not proved this.

Example 2:

Consider the following nonlinear regression function:

$Y(x) = \theta_1 sin\theta_2 x + \varepsilon = g(\theta_1, \theta_2; x) + \varepsilon$

.Following the same procedure as example 1 for the three different error distributions, we get the following estimating sequences. In each case $\theta_2^{(1)}$represents an arbitrary initial estimate of θ_2. We omit the details of the calculations.

Example 2 under (4.1)

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{1}{n}\frac{\left(Y_n - \theta_1^{(n)}(\theta_2^{(n)})sin\theta_2^{(n)}X_n\right)}{\left(\theta_1^{(n)}(\theta_2^{(n)})X_n cos\theta_2^{(n)}X_n\right)}, X_n \neq 0, n = 1,2,\ldots$$

$, \frac{\pi}{2} + v\pi, v = 0,1,2,\ldots$ and

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^n (sinx\theta_2^{(n)}X_i)^2}[\sum_{i=1}^n Y_i sin\theta_2^{(n)}X_i - \theta_1^{(n-1)}\sum_{i=1}^n (sin\theta_2^{(n)}X_i)^2],$$

$n = 1,2,\ldots; x_i \neq v\pi, v = 0,1,2,\ldots,$ where $\theta_1^{(1)} = Y_1/sin\theta_2^{(1)}X_i$

Example 2 under (4.2)

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{sign\left(Y_n - \theta_1^{(n)}(\theta_2^{(n)})sin\theta_2^{(n)}x_n\right)}{n(\theta_1^{(n)}(\theta_2^{(n)})x_n cos\theta_2^{(n)}x_n)}, n = 1,2,\ldots; x_n \neq 0$$

$, \frac{\pi}{2} + v\pi, v = 0,1,2,\ldots$ and

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n}(\sin\theta_2^{(n)}X_i)^2}\left[\sum_{i=1}^{n} Y_i \sin\theta_2^{(n)} X_i - \theta_1^{(n-1)}\sum_{i=1}^{n}(\sin\theta_2^{(n)}X_i)^2\right],$$

$n = 1,2,\dots;\quad x_i \neq v\pi, v = 0,1,2,\dots,$ where $\theta_1^{(1)} = Y_1/\sin\theta_2^{(1)}X_i$

Example 2 under (4.3)

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{\left(Y_n-\theta_1^{(n)}(\theta_2^{(n)})\sin\theta_2^{(n)}x_n\right)(r+3)\sqrt{\pi}[\int_0^{\frac{\pi}{2}}(\cos)^{(r-1)}\theta d\theta]^{-1}}{2n(\theta_1^{(n)}(\theta_2^{(n)})x_n\cos\theta_2^{(n)}x_n)(r+(Y_n-\theta_1^{(n)}(\theta_2^{(n)}\sin\theta_2^{(n)}x_n)^2)\Gamma\left(\frac{r+1}{2}\right)} \qquad (4.6)$$

$x_n \neq 0, \frac{\pi}{2} + v\pi, v = 0,1,2,\dots n = 1,2,\dots, r \geq 1$

, and ,

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n}(\sin\theta_2^{(n)}x_i)^2}\left[\sum_{i=1}^{n} Y_i \sin\theta_2^{(n)} X_i - \theta_1^{(n-1)}\sum_{i=1}^{n}(\sin\theta_2^{(n)}X_i)^2\right], \qquad (4.7)$$

$n = 1,2,\dots;\ x_i \neq v\pi, v = 0,1,2,\dots,$ where

$\theta_1^{(1)} = Y_1/\sin\theta_2^{(1)}X_i$

Taking different cases of degree of freedom r , the integral in (4.6) can be shown to be $\frac{\pi}{2}$, 1, 2/3, 384/945 for r=1,2,4,10, respectively. The estimating sequence $\theta_2^{(n)}$ in each case has the form

$$\theta_2^{(n+1)} - \theta_2^{(n)} + \frac{\lambda_r\left(Y_n-\theta_1^{(n)}(\theta_2^{(n)})\sin\theta_2^{(n)}x_n\right)}{n(\theta_1^{(n)}(\theta_2^{(n)})x_n\cos\theta_2^{(n)}x_n)(r+(Y_n-\theta_1^{(n)}(\theta_2^{(n)})\sin\theta_2^{(n)}x_n\ )^2)}$$

, n = 1, 2, … ; $x_n \neq 0$, and $\frac{\pi}{2} + v\pi, v = 0,1,2$.

Where $\lambda_r = 4,5,7,13$ for r = 1, 2, 4, 10 respectively. In each case the estimating sequence $(\theta_2^{(n-1)})$ is given by (4.7)

Example 3

We consider the following example. A certain chemical reaction can be described by the nonlinear model:

$$Y^* = \exp\left\{-\theta_1 x_1 \exp\left[-\theta_2\left(\frac{1}{x_2}-\frac{1}{620}\right)\right]\right\} \times \varepsilon^*$$

, where θ_1 and θ_2 are parameters to be estimated, Y^* is the fraction of original material remaining, x_1 is the reaction time in minutes, and x_2 is the temperature in degrees Kelvin. Taking natural logarithm and putting $Y = logY^*$ and $= log\varepsilon^*$, we have

$$Y(x_1,x_2) = -\theta_1 x_1 \exp\left[-\theta_2\left(\frac{1}{x_2}-\frac{1}{620}\right)\right] + \varepsilon$$

, following the same procedure as in Example 1 for the three different error distributions, we get the following estimating sequences. In each case $\theta_2^{(1)}$ represents an arbitrary initial estimate of θ_2. We omit the details,

<u>**Example 3 under (4.1)**</u>

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{\left(Y_n+\theta_1^{(n)}\left(\theta_2^{(n)}\right)x_{n1}\exp\left(-\theta_2^{(n)}\left(\frac{1}{x_{n2}}-\frac{1}{620}\right)\right)\right)}{n(\theta_1^{(n)}\left(\theta_2^{(n)}\right)x_{n1}\left(\frac{1}{x_{n2}}-\frac{1}{620}\right)\exp\left(-\theta_2^{(n)}\left(\frac{1}{x_{n2}}-\frac{1}{620}\right)\right))}$$

, n = 1, 2, …;

$x_{n1}, x_{n2} \neq 0$, and

$$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n}(x_{i1}\exp\left(-\theta_2^{(n)}\left(\frac{1}{x_{i2}}-\frac{1}{620}\right)\right))^2}\left[\sum_{i=1}^{n} Y_i x_{i1}\exp\left(-\theta_2^{(n)}\left(\frac{1}{x_{i2}}-\frac{1}{620}\right)\right) + \theta_1^{(n-1)}\sum_{i=1}^{n}\left(x_{i1}\exp\left(-\theta_2^{(n)}\left(\frac{1}{x_{i2}}-\frac{1}{620}\right)\right)\right)^2\right]$$

, n = 1, 2, … ; $x_{n1}, x_{n2} \neq 0$,i=1,2,….,n

Where $\theta_1^{(1)} = \frac{-Y_1}{(x_{i1}\exp\left(-\theta_2^{(n)}\left(\frac{1}{x_{i2}}-\frac{1}{620}\right)\right))}$

Example 3 under (4.2)

$$\theta_2^{(n+1)} = \theta_2^{(n)} + \frac{sign\left(Y_n+\theta_1^{(n)}\left(\theta_2^{(n)}\right)x_{n1}\exp\left(-\theta_2^{(n)}\left(\frac{1}{x_{n2}}-\frac{1}{620}\right)\right)\right)}{n(\theta_1^{(n)}\left(\theta_2^{(n)}\right)x_{n1}\left(\frac{1}{x_{n2}}-\frac{1}{620}\right)\exp\left(-\theta_2^{(n)}\left(\frac{1}{x_{n2}}-\frac{1}{620}\right)\right))}$$

, n = 1, 2, …;

$x_{n1} \neq 0, x_{n2} \neq 0$, and

$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n}(x_{i1}\exp(-\theta_2^{(n)}(\frac{1}{x_{i2}}-\frac{1}{620})))^2}[\sum_{i=1}^{n} Y_i x_{i1}\exp(-\theta_2^{(n)}(\frac{1}{x_{i2}}-\frac{1}{620})) + \theta_1^{(n-1)}\sum_{i=1}^{n}(x_{i1}\exp(-\theta_2^{(n)}(\frac{1}{x_{i2}}-\frac{1}{620})))^2]$

, n = 1, 2, … ; $x_{n1}, x_{n2} \neq 0$,i=1,2,….,n

Where $\theta_1^{(1)} = \frac{-Y_1}{(x_{i1}\exp(-\theta_2^{(n)}(\frac{1}{x_{i2}}-\frac{1}{620})))}$

Example 3 under (4.3)

$\theta_2^{(n+1)} = \theta_2^{(n)} + \{\frac{(r+3)\sqrt{\pi}\Gamma(\frac{r}{2})(Y_n+\theta_1^{(n)}(\theta_2^{(n)})x_{n1}\exp(-\theta_2^{(n)}(\frac{1}{x_{n2}}-\frac{1}{620})))}{2n(\theta_1^{(n)}(\theta_2^{(n)})x_{n1}(\frac{1}{x_{n2}}-\frac{1}{620})\exp(-\theta_2^{(n)}(\frac{1}{x_{n2}}-\frac{1}{620})))\Gamma(\frac{r+1}{2})}\}$

$\times \{\frac{[\int_0^{\frac{\pi}{2}}(\cos)^{(r+1)}\theta d\theta]^{-1}}{(r+(Y_n+\theta_1^{(n)}(\theta_2^{(n)})x_{n1}\exp(-\theta_2^{(n)}(\frac{1}{x_{n2}}-\frac{1}{620})))^2}\}$ (4.8)

, n = 1, 2, … ; $x_{n1} \neq 0, x_{n2} \neq 0, r \geq 1$, and

$\theta_1^{(n)} = \theta_1^{(n-1)} + \frac{1}{\sum_{i=1}^{n}(x_{i1}\exp(-\theta_2^{(n)}(\frac{1}{x_{i2}}-\frac{1}{620})))^2}[\sum_{i=1}^{n} Y_i x_{i1}\exp(-\theta_2^{(n)}(\frac{1}{x_{i2}}-\frac{1}{620})) + \theta_1^{(n-1)}\sum_{i=1}^{n}(x_{i1}\exp(-\theta_2^{(n)}(\frac{1}{x_{i2}}-\frac{1}{620})))^2]$

, (4.9)

n = 1, 2, … ; $x_{n1}, x_{n2} \neq 0$,i=1,2,….,n , Where $\theta_1^{(1)} = \frac{-Y_1}{(x_{i1}\exp(-\theta_2^{(n)}(\frac{1}{x_{i2}}-\frac{1}{620})))}$

Taking different cases of degree of freedom r , the integral in (4.6) can be shown to be $\frac{\pi}{2}$, 1, 2/3, 384/945 for r=1,2,4,10, respectively. The estimating sequence $\theta_2^{(n)}$ in each case has the form

$\theta_2^{(n+1)} = \theta_2^{(n)} + \{\frac{\lambda_r(Y_n+\theta_1^{(n)}(\theta_2^{(n)})x_{n1}\exp(-\theta_2^{(n)}(\frac{1}{x_{n2}}-\frac{1}{620})))}{n(\theta_1^{(n)}(\theta_2^{(n)})x_{n1}(\frac{1}{x_{n2}}-\frac{1}{620})\exp(-\theta_2^{(n)}(\frac{1}{x_{n2}}-\frac{1}{620})))\Gamma(\frac{r+1}{2})}\}$

$\times \{\frac{1}{(r+(Y_n+\theta_1^{(n)}(\theta_2^{(n)})x_{n1}\exp(-\theta_2^{(n)}(\frac{1}{x_{n2}}-\frac{1}{620})))^2}\}$

n = 1, 2, … ; $x_{n1}, x_{n2} \neq 0$,i=1,2,….,n

Where $\lambda_r = 4,5,7,13 \; for \; r = 1,2,4,10$ respectively. In each case the estimating sequence $\theta_1^{(n-1)}$ is given by (4.9)

5- Numerical Solution Using The Lawton & Syivestre Procedure

Lawton & Syivestre [18] considered the special case when the model has a linear and nonlinear component see equation(1.4). They introduce a modification based on the idea of reducing the number of parameters that must be estimated by the iterative methods. For a sample $Y_1, Y_2, \ldots\ldots Y_n$, the linear parameters are estimated at each stage by ordinary least squares and the estimates are substituted into (1.4). We now discuss three examples in some details. Computer program that would be required to obtain the numerical solutions are done. Details of a small simulation investigation of properties of the estimators obtained by Lawton & sylvestre procedure and SA-ILS procedure for the first model are given later in section (6)

Example 1:

We consider the model given in Example 1 , Section(4) that is $Y(x) = \theta_1 e^{\theta_2 x} + \varepsilon$.

Treating θ_2 as known, we then find θ_1 which minimizes the sum of squares:

$Q(\theta_1, \theta_2) = \sum_{i=1}^{n}(Y_i - \theta_1 e^{\theta_2 x_i})^2$, by Section (2), we have found the least square estimate of θ_1 given by : $\theta_1(\theta_2) = \frac{\sum_{i=1}^{n} Y_i e^{\theta_2 x_i}}{\sum_{i=1}^{n} e^{2\theta_2 x_i}}$, and the reduced "model" given by (1.6) as follows: $Y(x) = \theta_1(\theta_2)e^{\theta_2 x} + \varepsilon^*$ .where this "model", is treated as a nonlinear model with a single parameter θ_2, We have used a linearization method as an iterative

method for estimating the nonlinear parameter θ_2 . We will apply this method as follows:

Let

$$Z^{(j)} = \begin{bmatrix} x_1\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_1} \\ x_2\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_2} \\ \vdots \\ x_n\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_n} \end{bmatrix},$$

$$Y-g^{(j)}=\begin{bmatrix} Y_1-\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_1} \\ Y_2-\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_2} \\ \cdot \\ \cdot \\ Y_n-\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_n} \end{bmatrix}$$, j = 1,2,….

$(Z'^{(j)}Z^{(j)})^{-1}=$ $[\sum_{i=1}^{n}(x_i\theta_1(\theta_2^{(j)}))^2\exp(2\theta_2^{(j)}x_i)]^{-1}$ provided that $(Z'^{(j)}Z^{(j)})^{-1}$ exist, and let $(Z'^{(j)}(Y-g^{(j)})=$ $\sum_{i=1}^{n}[(x_i\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_i})(Y_i-\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_i})]$

, then , define an estimating sequence $(\theta_2^{(j)})$ by:

$$\theta_2^{(j+1)}=\theta_2^{(j)}+(Z'^{(j)}Z^{(j)})^{-1}(Z'^{(j)}(Y-g^{(j)}))$$

j=1,2,

i.e.,

$$\theta_2^{(j+1)}=\theta_2^{(j)}+[\sum_{i=1}^{n}(x_i\theta_1(\theta_2^{(j)}))^2\exp(2\theta_2^{(j)}x_i)]^{-1}[\sum_{i=1}^{n}[(x_i\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_i})\times(Y_i-\theta_1(\theta_2^{(j)})e^{\theta_2^{(j)}x_i})]]$$

,j= 1,2,…. Where $\theta_2^{(1)}$, is an initial estimate of θ_2.The above estimates , $\theta_2^{(j+1)}$, will be iteratively computed, in each iteration the "best" companion value of $\theta_1(\theta_2^{(j)})$, will be computed by the least square method.

Example 2

We consider the model given in Example 2 , Section (4) , that is $Y(x)=\theta_1 sin\theta_2 x+\varepsilon,$ we obtain $\theta_1(\theta_2)=\frac{\sum_{i=1}^{n}Y_i sin\theta_2 X_i}{\sum_{i=1}^{n}(sin\theta_2 X_i)^2}$, and , the reduced "model" will be $Y(x)=\theta_1(\theta_2)sin\vartheta_2 x+\varepsilon^*$,then, define an estimating sequence $(\theta_2^{(j)})$ by :

$$\theta_2^{(j+1)}=\theta_2^{(j)}+(Z'^{(j)}Z^{(j)})^{-1}(Z'^{(j)}(Y-g^{(j)}))$$

, j = 1,2,… provided that $(Z^{(j)}Z^{(j)})^{-1}$, exist , i.e.,

$$\theta_2^{(j+1)}=\theta_2^{(j)}+[\sum_{i=1}^{n}(x_i\theta_1(\theta_2^{(j)})\cos(\theta_2^{(j)}x_i))^2]^{-1}[\sum_{i=1}^{n}[(x_i\theta_1(\theta_2^{(j)})\cos(\theta_2^{(j)}x_i))\times(Y_i-\vartheta_1(\theta_2^{(j)})\sin(\theta_2^{(j)}x_i))]]$$

,j= 1,2,…Where $\theta_2^{(1)}$, is an initial estimate of θ_2.

Example 3

We consider the model given in Example 3 , Section (4), that is

$$Y(x_1,x_2)=-\theta_1 x_1\exp\left[-\theta_2\left(\frac{1}{x_2}-\frac{1}{620}\right)\right]+\varepsilon$$

, we obtain

$$\theta_1(\theta_2)=\frac{-\sum_{i=1}^{n}Y_i x_{i1}\exp(-\theta_2(\frac{1}{x_{i2}}-\frac{1}{620}))}{\sum_{i=1}^{n}(x_{i1}\exp\left(-\theta_2(\frac{1}{x_{i2}}-\frac{1}{620})\right))^2}$$

, and the reduced "model" will be

$$Y(x_1,x_2)=-\theta_1(\theta_2)x_1\exp\left[-\theta_2\left(\frac{1}{x_2}-\frac{1}{620}\right)\right]+\varepsilon^*$$

, then, define an estimating sequence $(\theta_2^{(j)})$by:

$$\theta_2^{(j+1)}=\theta_2^{(j)}+(Z'^{(j)}Z^{(j)})^{-1}(Z'^{(j)}(Y-g^{(j)}))$$

j = 1,2,… provided that $(Z'^{(j)}Z^{(j)})^{-1}$, exist , i.e.,

$$\theta_2^{(j+1)}=\theta_2^{(j)}+\left\{\frac{\sum_{i=1}^{n}(\theta_1(\theta_2^{(j)})x_{i1}(\frac{1}{x_{i2}}-\frac{1}{620})\exp(-\theta_2^{(j)}(\frac{1}{x_{i2}}-\frac{1}{620}))}{\sum_{i=1}^{n}\left((\theta_1(\theta_2^{(j)})x_{i1}(\frac{1}{x_{i2}}-\frac{1}{620})\exp\left[-\theta_2^{(j)}(\frac{1}{x_{i2}}-\frac{1}{620}\right]\right)^2}\right\}$$

$$\times\{\left(Y_i+\left(\theta_1(\theta_2^{(j)})x_{i1}\exp(-\theta_2^{(j)}(\frac{1}{x_{i2}}-\frac{1}{620}))\right)\right)\}$$

, j = 1,2,...

6- A Simulation Study

we report the finding of a small scale simulation study to compare the properties of the SA-ILS procedure and fixed sample size Lawton & Sylivestre procedure. The model used in example 1 was considered with $(x)=\theta_1\exp(\theta_2 x)+\varepsilon$, where $\varepsilon\sim N(0,\sigma^2)$. Values of (θ_1,θ_2), are taken as (0.15, 0.65), (0.1, 0.7), (0.25, 0.6) , and values of $\sigma^2=1,2,4$, were used to give markedly different pattern for the means and variances. For the fixed sample size procedure , a sample size n=10 was used. The residuals ε were generated using the random normal deviate generator available in the GLIM statistical computation system. It was decided that since this was only a preliminary study, the run size would be restricted to 10. Extreme caution must therefore be used in using the results to compare the properties of the two procedures. However, the results of this study indicate that for large sample size, the SA-ILS procedure would, generally speaking perform much better than the Lawton and Sylvestre procedure. The properties which we will be interested in for the two procedures are:

(1) $Mean = \theta_1 = \frac{1}{N}(\theta_{11} + \theta_{12} + \cdots + \theta_{1N}),$ $\theta_2 = \frac{1}{N}(\theta_{21} + \theta_{22} + \cdots + \theta_{2N})$;

(2)$Variance = S^2(\theta_1) = \frac{1}{N}\sum_{i=1}^{N}(\theta_{1i} - \theta_1)^2$, $S^2(\theta_2) = \frac{1}{N}\sum_{i=1}^{N}(\theta_{2i} - \theta_2)^2$;

(3) $SE(\theta_1) = \frac{1}{N}\sum_{i=1}^{N}(\vartheta_{1i} - \theta_1)^2$, $MSE(\theta_2) = \frac{1}{N}\sum_{i=1}^{N}(\theta_{2i} - \theta_2)^2$,

And the number of observations n_0 for SA-ILS procedure. The tables (1,2,3,4) give the estimates for θ_1 ,and θ_2 and compare their moment properties for the Lawton and Sylvestre , and SA-ILS procedures and give the sample numbers of observations required for the SA-ILS procedure. The values of the regressor variable were taken as 1(1)10.

7- Discussion And Conclusions

The following tentative conclusions can be made from the study

1. The sequential SA-ILS procedure required fewer observations than the fixed sample size procedure (n=10) in all cases , the number of observations required ranging from $n_o = 2 \; to \; n_o = 9$, and the average number of observations n_o varying between 4.7 to 7.1 . The advantage in reducing the number of observations was greater than the initial approximation $\vartheta_2^{(1)}$, was close to the true value. The results indicate that the SA-ILS procedure will lead to a real reduction in the number of observations required.
2. Different initial values for the SA-ILS procedure of course provided different for a given data set , and the estimates for a given model sometimes showed markedly different bias, variance and MSE properties. In general, as he initial value deviates from the true value, the variances and MSE values increase.
3. For fixed (θ_1, θ_2),and starting value $\theta_2^{(1)}$ for the SA-ILS procedure, increases in σ^2 led to estimates with increasing variance and MSE values, in nearly all cases.
4. Comparing the Lawton and Syivestre method with the SA-ILS procedure, it is seen that the biases of the estimates are larger for the first procedure. There was no clear pattern to distinguish the two procedures with regard to variances and MSE's.
5. From our simulation investigation we have noticed that the choice of initial starting value for the SA-ILS procedure is important, since a choice of value close to true value improves the behavior of the resulting estimators. It might therefore be useful to adopt a two-stage procedure combining both techniques in which a "small" fixed sample size n^* is selected and preliminary estimates θ_1^*, θ_2^* of θ_1, θ_2 made using the Lawton and sylivestre procedure. Then the sequential SA-ILS procedure could be used to generate further observations sequentially using θ_2^* as the initial value. This modification is not examined further in this study.

Table .1 Model : $Y_i = \theta_1 \exp(\theta_2 X_i) + \varepsilon_i$, where $(\varepsilon_i) \sim IN(0,1)$, with $\theta_1 = 0.15, \theta_2 = 0.65$

Lawton/Sylvestre Procedure			Stochastic Approximation - Iterative Least Squares Procedure																	
			$\hat\theta_2^{(1)} = 0.55$			$\hat\theta_2^{(1)} = 0.65$			$\hat\theta_2^{(1)} = 0.7$			$\hat\theta_2^{(1)} = 0.75$			$\hat\theta_2^{(1)} = 0.8$			$\hat\theta_2^{(1)} = 0.9$		
Run No.	$\hat\theta_1$	$\hat\theta_2$	$\hat\theta_1$	$\hat\theta_2$	n_o	$\hat\theta_1$	$\hat\theta_2$	n_o	$\hat\theta_1$	$\hat\theta_2$	n_o	$\hat\theta_1$	$\hat\theta_2$	n_o	$\hat\theta_1$	$\hat\theta_2$	n_o	$\hat\theta_1$	$\hat\theta_2$	n_o
1	0.1379	0.6587	0.1534	0.5057	4	0.1111	0.5809	4	9.395×10^{-2}	0.6316	4	7.911×10^{-2}	0.6750	4	9.207×10^{-2}	0.7224	7	5.679×10^{-2}	0.8101	6
2	0.1645	0.6390	0.1915	0.7484	3	0.1606	0.8113	3	0.1450	0.8456	3	5.499×10^{-2}	0.8672	5	2.170×10^{-2}	0.9006	8	1.173×10^{-2}	0.9784	8
3	0.1813	0.6298	0.3429	0.4702	6	0.2013	0.5710	6	0.1526	0.6197	6	0.1151	0.6682	6	8.636×10^{-2}	0.7172	6	4.806×10^{-2}	0.8167	6
4	0.1490	0.6500	0.1050	0.5595	5	0.1240	0.6454	5	0.1006	0.6896	5	8.118×10^{-2}	0.7347	5	6.017×10^{-2}	0.7797	7	3.210×10^{-2}	0.8731	7
5	0.1803	0.6303	0.3665	0.4466	5	0.2394	0.5414	5	0.1915	0.5900	5	0.1523	0.6393	5	0.1204	0.6891	5	6.026×10^{-2}	0.7887	7
6	0.1835	0.6290	0.1662	0.6003	5	0.1197	0.6715	5	0.1004	0.7091	5	8.354×10^{-2}	0.7480	5	6.908×10^{-2}	0.7879	6	4.458×10^{-2}	0.8703	6
7	0.1801	0.6313	0.4050	0.4927	4	0.2853	0.5866	4	0.1879	0.6200	7	0.1368	0.6754	7	0.1123	0.7243	6	6.350×10^{-2}	0.8225	6
8	0.1397	0.6590	0.0266	1.0990	4	0.0250	1.1150	4	2.370×10^{-2}	1.1300	4	2.216×10^{-2}	1.1470	4	2.040×10^{-2}	1.1680	4	1.700×10^{-2}	1.2160	4
9	0.1489	0.6500	0.1412	0.6747	6	0.1028	0.7493	5	8.532×10^{-2}	0.7884	5	7.033×10^{-2}	0.8285	5	2.707×10^{-2}	0.8655	8	1.391×10^{-2}	0.9498	8
10	0.1341	0.6610	0.1859	0.5814	6	0.1494	0.6405	7	0.1190	0.6745	7	9.300×10^{-2}	0.7112	5	7.147×10^{-2}	0.7502	5	4.051×10^{-2}	0.8335	7
(i)	0.1599	0.6439	0.2164	0.6187	4.8	0.1519	0.6922	4.8	0.1201	0.7307	5.1	8.805×10^{-2}	0.7695	5.1	6.811×10^{-2}	0.8105	6.2	3.884×10^{-2}	0.8959	6.5
(ii)	3.6×10^{-4}	1.6×10^{-4}	1.3×10^{-2}	3.3×10^{-2}	1.0	5.0×10^{-3}	2.6×10^{-2}	1.2	2.337×10^{-3}	2.3×10^{-2}	1.5	1.315×10^{-3}	2.1×10^{-2}	0.7	1.178×10^{-3}	1.8×10^{-2}	1.6	3.389×10^{-4}	1.5×10^{-2}	1.3
(iii)	4.7×10^{-4}	1.8×10^{-4}	1.7×10^{-2}	3.4×10^{-2}		5.0×10^{-3}	2.8×10^{-2}		1.232×10^{-3}	3.0×10^{-2}		5.055×10^{-3}	3.5×10^{-2}		7.887×10^{-3}	4.4×10^{-2}		1.270×10^{-2}	7.5×10^{-2}	
(iv)	9.9×10^{-3}	-6.1×10^{-3}	5.6×10^{-2}	-3.1×10^{-2}		1.9×10^{-3}	4.2×10^{-2}		-2.99×10^{-2}	0.1×10^{-2}		-6.115×10^{-2}	0.1195		-8.189×10^{-2}	0.1605		-0.1112	0.2459	

(i) Mean , (ii) Variance , (iii) Mean Square Error , (iv) Bias.

Table 2

Model : $Y_i = \theta_1 \exp(\theta_2 X_i) + \varepsilon_i$, where $(\varepsilon_i) \sim IN(0,2)$, with $\theta_1 = 0.15, \theta_2 = 0.65$

Lawton/Sylvestre Procedure			Stochastic Approximation - Iterative Least Squares Procedure																	
			$\hat{\theta}_2^{(1)} = 0.55$			$\hat{\theta}_2^{(1)} = 0.65$			$\hat{\theta}_2^{(1)} = 0.7$			$\hat{\theta}_2^{(1)} = 0.75$			$\hat{\theta}_2^{(1)} = 0.8$			$\hat{\theta}_2^{(1)} = 0.9$		
Run No.	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o
1	0.1328	0.6625	0.4480	0.4331	6	0.1216	0.4846	4	0.3453	0.5360	8	0.1558	0.5634	8	0.2117	0.6018	8	0.1369	0.6629	7
2	0.1715	0.6352	6.8×10^{-2}	0.7688	7	4.1×10^{-2}	0.8223	8	3.2×10^{-2}	0.8530	8	2.5×10^{-2}	0.8857	8	1.9×10^{-2}	0.9204	8	1.0×10^{-2}	0.9944	8
3	0.1968	0.6211	0.3601	0.4597	6	0.3261	0.5581	8	0.1570	0.6026	6	0.1178	0.6515	6	8.8×10^{-2}	0.7009	6	4.9×10^{-2}	0.8011	6
4	0.1488	0.6499	0.2024	0.5351	6	0.1271	0.6183	6	1.0×10^{-2}	0.6613	6	7.8×10^{-2}	0.7051	6	6.0×10^{-2}	0.7496	6	3.9×10^{-2}	0.8409	7
5	0.1948	0.6220	0.5071	0.4159	6	0.2751	0.5077	5	0.2197	0.5570	5	0.1786	0.6075	6	0.1347	0.6574	6	7.5×10^{-2}	0.7588	6
6	0.1996	0.6202	1.7×10^{-2}	1.0370	6	1.5×10^{-2}	1.0630	6	1.3×10^{-2}	1.0810	6	1.2×10^{-2}	1.1000	6	1.0×10^{-2}	1.1220	6	7.7×10^{-2}	1.1370	6
7	0.1948	0.6233	0.4847	0.4000	4	0.3399	0.5747	4	0.2132	0.6137	7	0.1549	0.6615	7	0.1117	0.7100	7	7.2×10^{-2}	0.8099	6
8	0.1352	0.6631	0.1010	0.7108	5	7.6×10^{-2}	0.7731	5	6.3×10^{-2}	0.8068	6	5.1×10^{-2}	0.8421	6	4.1×10^{-2}	0.8788	6	2.6×10^{-2}	0.9564	6
9	0.1483	0.6501	0.1574	0.6611	6	0.1141	0.7352	5	9.5×10^{-2}	0.7742	5	7.8×10^{-2}	0.8144	5	6.4×10^{-2}	0.8557	5	1.5×10^{-2}	0.9362	8
10	0.1274	0.6661	0.1607	0.5928	6	0.1486	0.6369	7	0.1226	0.6657	7	9.9×10^{-2}	0.6982	7	7.8×10^{-2}	0.7335	7	4.6×10^{-2}	0.8108	7
(i)	0.1650	0.6414	0.2507	0.6094	5.8	0.1583	0.6774	5.8	0.1360	0.7152	6.4	9.5×10^{-2}	0.7529	6.5	8.2×10^{-2}	0.7930	6.5	5.5×10^{-2}	0.8744	6.7
(ii)	7.9×10^{-4}	3.3×10^{-4}	3.0×10^{-2}	3.4×10^{-2}	0.6	1.2×10^{-2}	2.8×10^{-2}	2.0	9.2×10^{-3}	2.5×10^{-2}	1.0	2.9×10^{-3}	2.3×10^{-2}	0.9	3.2×10^{-3}	2.1×10^{-2}	0.9	1.3×10^{-3}	1.9×10^{-2}	0.6
(iii)	1.0×10^{-3}	3.4×10^{-4}	4.0×10^{-2}	3.5×10^{-2}		1.2×10^{-2}	2.9×10^{-2}		9.4×10^{-3}	2.9×10^{-2}		6.0×10^{-3}	3.4×10^{-2}		7.9×10^{-3}	4.2×10^{-2}		1.0×10^{-2}	5.9×10^{-2}	
(iv)	1.5×10^{-2}	-8.6×10^{-3}	0.1007	-4.1×10^{-2}		8.3×10^{-3}	2.7×10^{-2}		-1.4×10^{-2}	6.5×10^{-2}		-5.5×10^{-2}	0.1029		-6.8×10^{-2}	0.1430		-9.5×10^{-2}	0.2244	

(i) Mean , (ii) Variance , (iii) Mean Square Error , (iv) Bias.

Table . . 3 Model : $Y_i = \theta_1 \exp(\theta_2 X_i) + \varepsilon_i$, where $(\varepsilon_i) \sim IN(0,4)$, with $\theta_1 = 0.15, \theta_2 = 0.65$

Lawton/Sylvestre Procedure			Stochastic Approximation - Iterative Least Squares Procedure																	
			$\hat{\theta}_2^{(1)} = 0.55$			$\hat{\theta}_2^{(1)} = 0.65$			$\hat{\theta}_2^{(1)} = 0.7$			$\hat{\theta}_2^{(1)} = 0.75$			$\hat{\theta}_2^{(1)} = 0.8$			$\hat{\theta}_2^{(1)} = 0.9$		
Run No.	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o
1	0.1254	0.6684	0.3005	0.1288	4	0.3371	1.0×10^{-2}	4	0.4048	5.2×10^{-2}	4	0.5577	-4.0×10^{-2}	4	2.1935	5.0×10^{-2}	6	0.4860	0.9527	2
2	01821	0.6287	7.7×10^{-2}	0.8063	5	6.2×10^{-2}	0.8524	5	5.4×10^{-2}	0.8800	5	4.6×10^{-2}	0.9102	5	3.9×10^{-2}	0.9427	5	9.4×10^{-3}	1.0090	8
3	0.2206	0.6089	0.3816	0.4357	6	0.2176	0.5301	6	0.1632	0.5781	6	0.1220	0.6266	6	9.1×10^{-2}	0.6754	6	6.4×10^{-2}	0.7750	8
4	0.1490	0.6493	0.3420	0.4031	6	0.2936	0.4321	6	0.2979	0.4293	6	0.3462	0.4008	6	1.4720	0.3332	8	0.1164	0.9049	3
5	0.2168	0.6105	0.6421	0.3647	6	0.3991	0.4559	6	0.3101	0.5027	6	0.2391	0.5500	6	0.1832	0.5978	6	0.1061	0.6943	6
6	0.2242	0.6081	2.6×10^{-2}	0.9737	6	2.2×10^{-2}	1.0040	6	1.9×10^{-2}	1.0250	6	1.7×10^{-2}	1.0480	6	1.4×10^{-2}	1.0740	6	1.0×10^{-2}	1.1329	6
7	0.2172	0.6121	0.5976	0.4675	4	0.3379	0.5505	7	0.2471	0.5979	7	0.1792	0.6461	7	0.1290	0.6949	7	8.4×10^{-2}	0.7955	6
8	0.1288	0.6690	9.7×10^{-2}	0.7086	5	7.2×10^{-2}	0.7744	5	6.1×10^{-2}	0.8098	5	5.1×10^{-2}	0.8466	6	4.1×10^{-2}	0.8845	6	2.5×10^{-2}	0.9640	6
9	0.1468	0.6507	0.1487	0.6786	6	0.1003	0.7473	6	9.5×10^{-2}	0.7854	5	7.9×10^{-2}	0.8240	5	6.5×10^{-2}	0.8639	5	4.3×10^{-2}	0.9470	5
10	0.1183	0.6735	2.3×10^{-5}	1.8960	7	3.2×10^{-5}	1.8510	7	4.7×10^{-5}	1.7930	7	7.9×10^{-5}	1.7190	7	1.4×10^{-5}	1.6410	7	3.6×10^{-4}	1.5050	7
(i)	0.1729	0.6379	0.2636	0.6863	5.5	0.1841	0.7298	5.8	0.1651	0.7453	5.7	0.1636	0.7532	5.8	0.4228	0.7758	6.2	9.5×10^{-2}	0.9680	5.7
(ii)	1.7×10^{-3}	6.7×10^{-4}	4.7×10^{-2}	0.2166	0.9	2.0×10^{-2}	0.1982	0.8	1.8×10^{-2}	0.1889	0.8	2.8×10^{-2}	0.1880	0.8	0.5257	0.1656	0.8	1.9×10^{-2}	4.6×10^{-2}	3.4
(iii)	2.3×10^{-3}	8.4×10^{-4}	6.0×10^{-2}	0.2179		2.1×10^{-2}	0.2046		1.8×10^{-2}	0.1980		2.8×10^{-2}	0.1985		0.6001	0.1814		2.2×10^{-2}	0.1476	
(iv)	2.3×10^{-2}	-1.2×10^{-2}	0.1136	3.6×10^{-2}		3.4×10^{-2}	8.0×10^{-2}		1.5×10^{-2}	9.5×10^{-2}		1.4×10^{-2}	0.1032		0.2728	0.1258		-5.6×10^{-2}	0.3180	

(i) Mean , (ii) Variance , (iii) Mean Square Error , (iv) Bias.

Table [illegible].4 Model : $Y_1 = \theta_1 \exp(\theta_2 X_1) + \varepsilon_i$, where $(\varepsilon_i) \sim IN(0,1)$, with $\theta_1 = 0.1$, $\theta_2 = 0.7$

Lawton/Sylvestre Procedure			Stochastic Approximation - Iterative Least Squares Procedure																	
			$\hat{\theta}_2^{(1)} = 0.5$			$\hat{\theta}_2^{(1)} = 0.6$			$\hat{\theta}_2^{(1)} = 0.7$			$\hat{\theta}_2^{(1)} = 0.8$			$\hat{\theta}_2^{(1)} = 0.9$			$\hat{\theta}_2^{(1)} = 1.0$		
Run No.	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o	$\hat{\theta}_1$	$\hat{\theta}_2$	n_o
1	8.4×10^{-2}	0.7167	-3.6×10^{-2}	-0.5016	4	-8.4×10^{-3}	-1.5650	4	1.8×10^{-6}	2.9030	5	-9.5×10^{-3}	0.1551	3	-0.1708	-0.4138	4	-4.3×10^{-2}	0.8188	2
2	7.8×10^{-2}	0.7249	0.2899	0.4701	5	0.1940	0.5596	6	0.1149	0.6515	6	6.1×10^{-2}	0.7457	7	3.2×10^{-2}	0.8423	7	1.4×10^{-2}	0.9404	8
3	9.3×10^{-2}	0.7089	0.1319	0.6435	7	0.3072	0.7413	3	6.5×10^{-2}	0.7853	6	4.1×10^{-2}	0.8643	6	2.5×10^{-2}	0.9476	6	1.5×10^{-2}	1.0340	6
4	9.1×10^{-2}	0.7103	0.2678	0.3892	4	0.2465	0.4153	4	0.2135	0.4593	4	0.1817	0.5140	5	0.1416	0.5734	6	0.1002	0.6345	6
5	0.1006	0.7000	1.3×10^{-2}	0.8245	7	3.5×10^{-2}	0.8543	7	2.2×10^{-2}	0.8978	8	1.4×10^{-2}	0.9525	8	8.7×10^{-3}	1.0150	8	5.0×10^{-3}	1.0850	8
6	8.9×10^{-2}	0.7131	0.2630	0.5398	5	0.1873	0.6141	5	0.1047	0.6906	7	6.0×10^{-2}	0.7731	7	3.3×10^{-2}	0.8594	7	1.8×10^{-2}	0.9485	7
7	0.1072	0.6930	1.2×10^{-7}	3.3500	5	2.7×10^{-3}	1.3118	6	1.3×10^{-2}	1.0480	6	1.7×10^{-2}	1.0012	6	1.5×10^{-2}	1.0230	6	1.1×10^{-2}	1.0749	6
8	7.8×10^{-2}	0.7259	-0.1437	1.4×10^{-2}	2	5.0×10^{-4}	1.7750	5	3.1×10^{-2}	0.9324	5	7.0×10^{-2}	0.7244	6	0.1129	0.6589	6	0.1199	0.6402	6
9	0.1058	0.6957	0.1253	0.6711	6	8.5×10^{-2}	0.7375	6	4.8×10^{-2}	0.8090	7	2.4×10^{-2}	0.8857	8	1.3×10^{-2}	0.9671	8	6.5×10^{-3}	1.0520	8
10	0.1027	0.6976	0.2265	0.4943	5	0.1504	0.5806	5	9.7×10^{-2}	0.6705	5	6.2×10^{-2}	0.7631	6	3.6×10^{-2}	0.8575	6	2.0×10^{-2}	0.9539	6
(i)	9.3×10^{-2}	0.7086	0.1168	0.6933	5	0.1201	0.6025	5.1	7.1×10^{-2}	0.9047	5.9	5.3×10^{-2}	0.7379	6.2	3.2×10^{-2}	0.7330	6.4	2.7×10^{-2}	0.9190	6.3
(ii)	1.1×10^{-2}	1.3×10^{-4}	2.0×10^{-2}	0.9176	2	1.2×10^{-2}	0.6714	1.3	3.7×10^{-3}	0.4337	1.3	2.5×10^{-3}	5.5×10^{-2}	2.0	6.3×10^{-3}	0.1654	1.2	2.1×10^{-3}	2.5×10^{-2}	2.8
(iii)	1.6×10^{-4}	2.1×10^{-4}	2.0×10^{-2}	0.9176		1.2×10^{-2}	0.6809		4.5×10^{-3}	0.5148		4.7×10^{-3}	0.1297		1.1×10^{-2}	0.1666		7.4×10^{-3}	7.3×10^{-2}	
(iv)	-7.2×10^{-3}	8.6×10^{-3}	1.7×10^{-2}	6.7×10^{-3}		2.0×10^{-2}	-9.8×10^{-2}		-2.9×10^{-2}	0.2047		-4.7×10^{-2}	3.8×10^{-2}		-6.8×10^{-2}	3.3×10^{-2}		-7.3×10^{-2}	0.2190	

(i) Mean , (ii) Variance , (iii) Mean Square Error , (iv) Bias.

References

1. A, Sami N **(1999**),*Transformation of observation in stochastic approximation ,Ann.Statist.*1,6,p 1158-1174
2. A, Sami N (**2002**).On optimal multidimensional stochastic approximation; A unified approach 1 ,The proceeding of math, and computer science conference, UNESCO and Alexandria Univ. ,Nov. **2002**. p 393-413
3. A, Sami N (**2007)** .On optimal multidimensional stochastic approximation procedures, Florida State university statistical Report no. M-603,October,(**2007)** ,submitted to Communications in Statistics
4. A, Sami N (**2009**) .On estimation Via optimal stochastic approximation procedures, Statistical Theory and Data Analysis, K. Matusita (Editor) Elserier Science Publishers B.V.(North Holland)
5. Albert, A.e. and Gardner,L.A (2000) . Stochastic approximation and nonlinear regression,Reaearch Monograph No.42,M.I.T. Press.
6. Anber, D.(**2002**). On optimal estimation method using stochastic approximation procedures,Ann.statist.1,6,p 1175-1184
7. Bartle,R.G. (2001) . The Elements of real Anaysis, Johin Wiley &Sons,Inc.,New York
8. Blum,J.R. (**1994a**) . Approximation methods which converge with probability one,*Ann.Math .statist***.**25,p 362-386
9. Bium, J.B**.** **(1994b).** Multidimnsional stochastic approximation methods**,** *Ann .math,Statist***.**25,737-744
10. Chung, K.L.**(1997).** A course in Probability Theory**,** Academic Press,Inc.,New York
11. Draper, N.R. and Smith, h**.(2004**). Applied Regression Analysis**,** johin Wiley & Sons,Inc.,New York
12. Fabian m V.(**1990**). Stochastic approximation methods, Czech.Math.J.10,p 123-159
13. Fabian,V.(**2000**). On asymptotic normality in stochastic approximation , *Ann, Math.Statist*.39,p 1327-1332
14. Fabian, V.(**1991**), Stochastic approximation, Proc. Symp. Optimizing Math.in Statistics (J.S.Rustagi,ed.) p 439-470 , Academic Press.
15. Gladyshev,E.G. (**1996**). On stochastic approximation,Theory of probability and its Applications X(**1996**) ,No.2, p 275-278
16. Graybill,F.A.(**1996**) Introduction to Matrices with approximations in Statistics,Wadsworth publishing Co.,Inc.,California
17. Kiefer,J.and Wolfowitz,J.(**2001**) . Stochastic estimation of the maximum of the maximum of a regression function, Ann,math, Statist. **23**, 462-466.
18. Lawton,W.H.and Sylvestre,E.A.(**1998**). Elimination of linear parameters in nonlinear regression,Teahnometrics,13,p 461-467,Discussion , p 477-481
19. Lawton,W.H.,Sylv evstre,E.A, and Maggio,M.S.(**1992**) . Self modeling nonlinear regression,technomatrics,14, p 513-532
20. McCullagh,p. and Neider,J,K.(**2008**). Generalazed Linear Models,Chapman-Hall-London
21. Neveison,M.V.and Has'minskil,R.Z.(**2006**).Stochastic approximation and recursive estimation, Transilation of mathematical Monograph 47 (American mathmetical socisty)

22. Robbins, b. and Monro,s.(**2001**).A stochastic approximation method, *Ann. math. Statist.12***,**p 400-407
23. Robbins,H.and siegmund,D**.(2005),**a convergence theorem for nonnegative almost supermatingales and some approximations,Proc.Symp. Optimizing Methods in Statistics (J.S.rustagi,ed.).p 233-257, Academic Press,New York