Journal of Al-Nahrain University Vol.16 (2), July, 2013, pp.202-209 Science

Asymptotic Properties of the most Generalized Optimal Stochastic Approximation Procedures

Ali H. Kashmar
Department of Computer Science, College of Science, Baghdad University.

Abstract

In this paper we consider the most general nonlinear regression model, $\mathrm{Y}(x) = \psi(\theta_{(1)}) g_1(\theta_{(2)}; x) + \varepsilon$, prove of the almost sure convergence, and asymptotic normality of the estimators for the nonlinear parameters, using the most general optimal stochastic approximation procedure. A procedure for constructing the general confidence intervals for the vector of nonlinear parameters is also developed; the most generalized nonlinear regression model is introduced. We establish asymptotic properties for the most generalized model.

1-Interduction

Consider the following nonlinear regression model: $Y(x) = g(\theta; x) + \varepsilon$ where $g: R^p \times R^r \to R$, with $R^p \ and \ R^r$, being Euclidean spaces, ε is an unobservable random error, with $E(\varepsilon) = 0$, $var(x) = \sigma^2; \sigma^2$ is a constant that may depend on x ; Y(x) is an observable random that can be observed at each level $x \in R^r$; and $\theta \in R^p$ is the parameter of interest. Based on observations $Y_1, Y_2, \dots, Y_n$, it has been known, [6],[10],via classical procedures, how to estimate $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_p)'$. Now our problem is to estimate θ sequentially by using optimal stochastic approximation methods [1],[2],[3],[4]. We shall incorporate the approach of eliminating linear parameters proposed by [9] in an iterative manner together with optimal stochastic approximation procedures to estimate θ sequentially. We shall also consider the general model $(Y(x) = g(\theta; x) + \varepsilon \quad)$ when the regression function ,as far as the parameters are concerned , is composed of linear and nonlinear regression functions, i.e., $g(\theta; x) = \psi((\theta_{(1)}) g_1(\theta_{(2)}; x)$, where $g_1(\theta_{(2)}; x)$, is a nonlinear regression function and $g_1: R^{p-q} \times R^r \to R^m$, with $\theta_{(2)} \in R^{p-q}$ and $\psi((\theta_{(1)})$ is a real valued linear function of $(\theta_{(1)})$. However $\psi((\theta_{(1)})$ may depend on x, and $\theta_{(1)} \in R^q$. The model: $\underset{\sim}{Y}(\underset{\sim}{x}) = g\left(\underset{\sim}{\theta}; \underset{\sim}{x}\right) + \underset{\sim}{\varepsilon}$ then takes the form $Y(x) = \psi((\theta_{(1)}) g_1(\theta_{(2)}; x) + \underset{\sim}{\varepsilon}$ .We study the most general stochastic approximation procedure to estimate $\theta_{(2)}$, given by:

$$\acute{\theta}_{\sim(2)}^{(n+1)} = \theta_{(2)}^{(n)} - \alpha_n \psi\left(\acute{\theta}_{\sim\ (1)}^{\ )\ (n)}\right) A_n h_n \quad , \quad n = 1,2, \dots.$$

with $\acute{\theta}_{\sim(2)}^{(1)}$ being an arbitrary random vector in R^{p-q} that will be independent of any future observations, α_n may be positive measurable functions of $(\acute{\theta}_{\sim(2)}^{(1)}; x_1, x_2 \dots . x_n; h_1, h_2 \dots, h_{n-1})$, A_n are (p-q x m) matrix-valued measurable functions of $(\acute{\theta}_{\sim(2)}^{(1)}; x_1, x_2 \dots, x_n; h_1, h_2 \dots . h_{n-1})$ and h_n are design vectors in R^m based on transforming the observations Y_n by a Borel measurable transformation, $h = (h^{(1)}, \dots \dots, h^{(n)})'$. We study the almost sure convergence as well as the asymptotic normality of the sequential estimating sequence $\acute{\theta}_{\sim(2)}^{(n)}$ generated by $(\acute{\theta}_{\sim(2)}^{(n+1)} = \acute{\theta}_{\sim(2)}^{(n)} - \alpha_n \psi\left(\acute{\theta}_{\sim(2)}^{(n)}\right) A_n h_n \quad , \quad n = 1,2, \dots.)$, under conditions on h and on the conditional distribution of the error random vectors $V_n = Y_n - E\left(Y_n \middle| \acute{\theta}_{\sim(2)}^{(1)}, \acute{\theta}_{\sim(2)}^{(2)} \ , \dots, \acute{\theta}_{\sim(2)}^{(n)}\right) = Y_n - H_n(\theta_{(2)}^{(n)})$, we prove the asymptotic normality of $n^{\alpha/2}(\acute{\theta}_{\sim(2)}^{(n)} - \theta_{(2)})$ for $0 < \alpha \leq 1$, which is given as follows $n^{\alpha/2}(\acute{\theta}_{\sim(2)}^{(n)} - \theta_{(2)})$ is normal with mean = 0 , and covariance matrix = $a^2 l_2^2 P \sum P$, where $\sum^{(i\,j)} = (P\, A {\sum}_{\circ}(h, f) AP)^{(i\,j)} (a\Lambda^{(ii)} + a\Lambda^{(jj)} - \beta)^{-1}$. We construct general confidence intervals for the vector of nonlinear parameters, $\theta_{(2)}$ which is given by $n^{\alpha/2}\left(\acute{\theta}_{\sim(2)}^{(n)}-\theta_{(2)}\right)(\acute{\theta}_{\sim(2)}^{(n)}-\theta_{(2)}) \leq al_2 Z_{1-\alpha}||\sqrt{P\sum P}||$

Consider the following nonlinear regression model:

$Y(x)=g(\theta;x)+\varepsilon$ ....(1.1)

Where: $R^P \times R^r \rightarrow R^m$, with $R^P, R^r, and\ R^m$ are Euclidean spaces, ε is an unobservable vector of random errors, with $E(\varepsilon)=0$, $var(\varepsilon)=I\ \sigma^2$; where I is an identity matrix , σ^2 is a constant that may depend on x; Y(x) is an observable response random vector at each level $x \in R^r; and\ \theta \in R^p$ is the vector of parameters of concern.

We shall consider the most general nonlinear regression model of (1.1), when the regressionfunction is composed of linear and nonlinear regression components, i.e.

$g(\theta;x)=\psi(\theta_{(1)})g_1(\theta_{(2)};x)$.

Where $g_1(\theta_{(2)};x)$ is a nonlinear regression function in $\theta_{(2)}$, with $\theta_{(2)} \epsilon R^{p-q}$ and $\psi(\theta_{(1)})$ is a real –valued linear function of $\theta_{(1)}$. However $\psi(\theta_{(1)})$ may depend on x and $\theta_{(1)} \epsilon R^q$. Let ε be distributed according to a distribution function F that admits a symmetric density function f whose gradient vector exists, and the information matrix, I(f), off is positive definite. The model (1.1) then takes the form:

$Y(x)=\psi\left(\acute{\theta}_{\sim(1)}\right)g_1\left(\acute{\theta}_{\sim(2)};x\right)+\varepsilon,$ ....(1.2)

where $\psi: R^q \rightarrow R, and\ g_1: R^{p-q}\times R^r \rightarrow R^m$,

with $x=(x_1 \dots, x_r)$,

$\theta_{(1)}=\left(\theta_{1,1},\theta_{1,2},\dots\theta_{1,q}\right)$, and

$\theta_{(2)}=\left(\theta_{2,1},\theta_{2,2},\dots\theta_{2,p-q}\right)$.

Let us first estimate $\theta_{(1)}$ estimated sequentially by applying an iterative least square procedure. [4] ,since $\theta_{(1)}$ is assumed to appear linearly in the model (1.2), then substitute the initial guess $\acute{\theta}_{\sim(2)}^{(1)}$ of $\theta_{(2)}$ into (1.2). Therefore, the estimating sequence $\acute{\theta}_{\sim(1)}^{(n)}$ is given by

$$\acute{\theta}_{\sim(1)}^{(n)}=\acute{\theta}_{\sim(1)}^{(n-1)} + [\left((G_{\theta_{(2)}^{(n)}}G_{\theta_{(2)}^{(n)}})^{-1}G_{\theta_{(2)}^{(n)}}Y_n - \theta_{(1)}^{(n-1)}\right)], n=1,2,\dots$$

Where $\acute{\theta}_{\sim(1)}^{(1)}=[G_{\theta_{(2)}^{(1)}}G_{\theta_{(2)}^{(1)}}]^{-1}G_{\theta_{(2)}^{(1)}}Y_1$, is an initial value for the sequence $(\acute{\theta}_{\sim(1)}^{(n)})$, and $\acute{\theta}_{\sim(2)}^{(1)}$ is an arbitrary initial value for the sequence $(\acute{\theta}_{\sim(2)}^{(n)})$.Now substitute $\acute{\theta}_{\sim(1)}$ into (1.2). The vector of linear parameters $\acute{\theta}_{\sim(1)}$,is automatically replaced by the best companion value $\acute{\theta}_{\sim(1)}$ which is a function of $\acute{\theta}_{\sim(2)}$ alone. One then obtains the reduced model given by:

$Y(x)=\psi\left(\acute{\theta}_{\sim(1)}\right)g_1\left(\acute{\theta}_{\sim(2)};x\right)+\varepsilon^*$..........(1.3)

Consider the most general stochastic approximation procedures given by:

$\acute{\theta}_{\sim(2)}^{(n+1)}=\acute{\theta}_{\sim(2)}^{(n)}-\alpha_n\psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right)A_n h_n,$

$n=1,2,\dots,$ ....(1.4)

Where $\acute{\theta}_{\sim(2)}^{(1)}$ is an arbitrary random vector in R^{p-q} that will be independent of any future observations, α_n be positive measurable functions of $\left(\acute{\theta}_{\sim(2)}^{(1)},x_1,x_2\dots..x_n;h_1,h_2\dots.h_{n-1}\right)$, A_n are (p-q x m) n matrix-valued measurable functions of $\left(\acute{\theta}_{\sim(2)}^{(1)},x_1,x_2\dots.x_n;h_1,h_2\dots.h_{n-1}\right)$, and h_n are design vectors in R^m based on transforming the observations Y_n by a Boral measurable transformation, $h=(h^{(1)},\dots\dots,h^{(m)})$

Our main objective is to study the almost sure convergence of $\acute{\theta}_{\sim(2)}^{(n)}$. weshall study the aasymptotic normality, of the sequential estimating sequence $(\acute{\theta}_{\sim(2)}^{(n)})$ generated by (1.4). Weshall construct confidence intervals for the vector of nonlinear parameters $\acute{\theta}_{\sim(2)}$.

2- Assumptions

The following assumptions are stated in this section to be called upon later in the sequel [5]

Assumption (2.1):

$\acute{\theta}_{\sim(2)}^{(1)}, \acute{\theta}_{\sim(2)}^{(2)}, \ldots..$ Are (p-q)-dimensional random vectors. Let $h: R^m \to R^m$ be a Boral measurable transformation such that for $n \in N, E(h(Y_n(\acute{\theta}_{\sim(2)}^{(n)})))$ exists. Let $Y_1, Y_2, \ldots\ldots;\ and\ h_1, h_2, \ldots.$ be-m-dimensional random vectors with $E\left(h_n \middle| \acute{\theta}_{\sim(2)}^{(1)}, \acute{\theta}_{\sim(2)}^{(2)}, \ldots \acute{\theta}_{\sim(2)}^{(n)}\right) = E\left(h(Y_n) \middle| \acute{\theta}_{\sim(2)}^{(1)}, \ldots \acute{\theta}_{\sim(2)}^{(n)}\right) = E\left(h(Y_n) \middle| \acute{\theta}_{\sim(2)}^{(n)}\right) = E\left(h\left(Y_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right)\right),$

For all $n \in N$.Moreover, let α_n be positive measurable functions of $\left(\acute{\theta}_{\sim(2)}^{(1)}, x_1, x_2 \ldots..x_n; h_1, h_2 \ldots h_{n-1}\right)$, and A_n be (p-q x m) matrix-valued measurable functions in$R^{(p-qxm)}$. Let $\psi\left(\acute{\theta}_{\sim(1)}\right)$be a linear function of$\acute{\theta}_{\sim(1)}; \acute{\theta}_{\sim(1)} \in R^q$;ψ(.) may depend on $x \in R^r, \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right)$is an iterative least squares estimate of$\psi\left(\acute{\theta}_{\sim(1)}\right)$, and, with an initial estimate $\acute{\theta}_{\sim(2)}^{(1)}$of $\acute{\theta}_{\sim(2)}$, let

$$\acute{\theta}_{\sim(2)}^{(n+1)} = \acute{\theta}_{\sim(2)}^{(n)} - \alpha_n \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n h_n\ , \quad n = 1,2 \ldots\ldots, \qquad \text{.......................(2.1)}$$

Moreover, let $H_{n,h}(.)$ and $H_n(.)$ be Boral measurable functions defined on $R^{p-q}\ into\ R^m, and\ \psi(.)$ be bounded Boral measurable function defined on $R^q\ into\ R$, $with\ F_1 = \sigma(\acute{\theta}_{\sim(2)}^{(1)},\ \ F_n = \sigma\left(\acute{\theta}_{\sim(2)}^{(1)}, x_1, x_2, \ldots\ldots x_n; h_1, h_2, \ldots.h_{n-1}; A_1, \ldots\ldots A_n\right),$ (The smallest-σ-field induced by the indicated functions), let

$$E_{F_n}(h_n) = H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right) \text{.............................(2.2)}$$

with $H_{n,h} = H_n$, where h = the identity.

Assumption (2.2)

The m-dimensional observations $Y_1, Y_2, \ldots$ satisfy $Y_n = H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right) + V_n$, $n = 1,2, \ldots$ where $V_n = Y_n - E\left(Y_n \middle| \acute{\theta}_{\sim(2)}^{(1)}, \acute{\theta}_{\sim(2)}^{(2)}, \ldots\ldots \acute{\theta}_{\sim(2)}^{(n)}\right) = Y_n - H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)$, are random vector errors. V_n, are $conditionally\left(given\ \left(\acute{\theta}_{\sim(2)}^{(1)}, \ldots\ldots \acute{\theta}_{\sim(2)}^{(n)}\right)\right)$ distributed with a distribution function F that admits a symmetric density function f whose gradient vector exists and the information matrix, I(f) , of f is positive definite .

Assumption (2.3)

Let$(t) = \left(u^{(1)}(t), \ldots\ldots.u^{(m)}(t)\right), t \in R^m$, be defined by

$$(U(t)^{(j)} = u^{(j)}(t)) = \int_{R^m} h^{(j)}(t+v) f(v) dv, j = 1,2, \ldots.., m$$

Which exists for all t $\in R^m$. Moreover, suppose that

$$\int_{R^m} [\,h^{(i)}(t+v) - u^{(i)}(t)]\,[h^{(j)}(t+v) - u^{(j)}(t) f(v) dv]$$

Exists and is finite for all $t \in R^m$, and i , j = 1,2,….m.

Assumption (2.4):

The transformation h satisfies: $\int_{R^m} h^{(i)}(v) f(v) dv = 0 \quad and \quad \int_{R^m} h^{(i)}(v) h^{(j)}(v) f(v) dv = \begin{cases} 0 & for\ i \neq j \\ \sigma_i^2(h,f) & for\ i = j \end{cases}$

Where $\sigma_i^2(h,f)$ is positive constant that depends on h and f;i;j=1,2,….,m. This assumption actuallystates that the components of the vector $h(V_n)$ are uncorrelated, i.e. $cov\left(h^{(i)}, h^{(j)}\right) = 0$.

3-Almost Sure Convergence of the Most General Stochastic Approximation Procedures

To establish almost sure convergence of the most general stochastic approximation procedure in (*Assumption2. 1*), we shall refer to Almost Sure convergence Theorem [11]

3.1-<u>(almost sure Convergence Theorem</u>)

Let $\theta_{(2)} \in R^{p-q}, \theta_{(2)}^{(n)} \in R^q$,and assumption (2. 1) hold with $A_n(p-q\ x\ m)$ matrix-valued F_n-measurable functions of $(\acute{\theta}_{\sim(2)}^{(1)}, x_1, x_2, \ldots.x_n; h_1, \ldots.h_{n-1})$.Let $\rho_n.v_n, l_n, and\ t_n$ be positive numbers. Then suppose for every $\acute{\theta}_{\sim(2)}^{(n)} \in R^{p-q}$, and every $n \geq 1,(< \acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)},$

$\psi(\acute{\theta}_{\sim(1)}^{(n)})\ A_n H_{n,h}(\acute{\theta}_{\sim(2)}^{(n)}) >) \geq 0;$ ....(3.1)

If $0 < \delta_1 < \delta_2 < \infty$, then, for every $n \geq 1$,

$inf_{\delta_1 \leq \|\theta_{(2)}^{(n)} - \theta_{(2)}\| \leq \delta_2} \left\| H_{n,h}(\theta_{(2)}^{(n)}) \right\| \geq \rho_n$ ......(3.2)

$inf_{\theta_{(2)}^{(n)}} \left\{ < \theta_{(2)}^{(n)} - \theta_{(2)}, \psi(\theta_{(2)}^{(n)} A_n H_{n,h}(\theta_{(2)}^{(n)}) > / (\psi\left(\theta_{(1)}^{(n)}\right) \left\| \theta_{(2)}^{(n)} - \theta_{(2)} \right\| \times \left\| H_{n,h}\left(\theta_{(2)}^{(n)}\right) \right\|) \right\} \geq v_n$ ....(3.3)

For each sequence $(\theta_{(2)}^{(n)})$ for which $^{sup}_{n}||\theta_{(2)}^{(n)}|| < \infty$. Suppose also that $\sum_{n=1}^{\infty} \alpha_n v_n l_n \infty \ \ and \ \ \sum_{n=1}^{\infty} \alpha_n v_n l_n \rho_n = \infty$, ...(3.4)

For every sequence $\acute{\theta}_{\sim(2)}^{(1)}, x_1, \ldots \ldots x_n, h_1, h_2, \ldots$. for which $^{sup}_{n}||\theta_{(2)}^{(n)}|| < \infty$,

$\left\| H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right) \right\| \leq \tau_n \left(\left\| \acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)} \right\| + 1 \right)\ ;\ \acute{\theta}_{\sim(2)}^{(n)} \in R^{p-q}, n \geq 1;$........(3.5)

$\psi(\acute{\theta}_{\sim(1)}^{(n)}) \geq l_n\ ;\ \acute{\theta}_{\sim(1)}^{(n)} \in R^q\ ,\ n \geq 1$........(3.6)

Furthermore, for every sequence $\acute{\theta}_{\sim(2)}^{(1)}, x_1, \ldots\ldots, x_n, h_1, h_{2,\ldots\ldots}$, let $\sum_{n=1}^{\infty} l_n^2 \alpha_n^2\, \tau_n^2 ({}_{(\theta_{(2)}^{(1)}, x_1, \ldots x_n; h_1, \ldots h_{n-1})}^{sup} ||A_n||^2) < \infty$ ....(3.7)

And $\sum_{n=1}^{\infty} l_n^2 \alpha_n^2 E_{F_n}[||h_n - H_{n,h}(\acute{\theta}_{\sim(2)}^{(n)}||^2]\ {}_{(\theta_{(2)}^{(1)}, x_1, \ldots x_n;; h_1, \ldots, h_{n-1})}^{sup} ||A_n||^2),$ $< \infty$ ….(3.8)

Then $\acute{\theta}_{\sim(2)}^{(n)} \to \acute{\theta}_{\sim(2)}$ almost surely as $\to \infty$. Before the proof we shall give some remarks on the above conditions.

Remarks:

(1) We use eq. (3.1), since $\left(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}\right)$ must be in the same direction of $A_n \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right)$, so that the regression function is bounded between two lines, also to insure that the angle and its cosine must be positive between two vectors $\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}$ and $\psi(\acute{\theta}_{\sim(1)}^{(n)}) A_n H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right)$.

(2) Eq. (3.2) allows $\left\| H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right) \right\| \to 0\ as\ \left| \left| \acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)} \right| \right| \to 0$, and eq. (3.3) is Cauchy-Schwartz inequality, it is concerned with the cosine of the angle between two vectors, it requires that the smallest point must be greater or a positive number, such that $^{sup}_{n}\left| \left| \acute{\theta}_{\sim(2)}^{(n)} \right| \right| < \infty$.

(3) Eq. (3.5) is bounded condition on $H_{n,h}(\acute{\theta}_{\sim(2)}^{(n)})$ by using Lipschitz condition. Eq. (3.6) is a bounded condition on $\psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right)$, from below.

(4) Eq. (3.7) implies $||A_n||$ multiplied by $\alpha_n^2, l_n^2, \tau_n^2$ is bounded, and eq. (3.8) mean that conditional variance of h_n multiplied by $\alpha_n^2, l_n^2,$ and ${}_{(\theta_{(2)}^{(1)}, x_1, \ldots x_n; h_1, \ldots h_{n-1})}^{sup} ||A_n||^2$ is bounded.

Proof:

Without loss of generality, let $\acute{\theta}_{\sim(2)} = 0$. By (2.1), we have

$\acute{\theta}_{\sim(2)}^{(n+1)} = \acute{\theta}_{\sim(2)}^{(n)} - \alpha_n \psi(\acute{\theta}_{\sim(1)}^{(n)} A_n h_n\ ,\ \ n = 1,2,$.and thus

$$\acute{\theta}_{\sim(2)}^{(n+1)} = \acute{\theta}_{\sim(2)}^{(n)} - \alpha_n \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n H_{n,h}(\acute{\theta}_{\sim(2)}^{(n)} + \alpha_n \psi\,(\acute{\theta}_{\sim(1)}^{(n)})\, A_n Z_n \quad \text{.........................(3.9)}$$

$where\ Z_n = -\left(h_n - H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right)\ with\ E_{F_n} Z_n = 0$. Thus, by using (3.5) and (3.6) , we get

$$||\acute{\theta}_{\sim(2)}^{(n+1)}||^2 = ||\acute{\theta}_{\sim(2)}^{(n)} - \alpha_n \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n H_{n.h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right) + \alpha_n \psi(\acute{\theta}_{\sim(1)}^{(n)}) A_n Z_n||^2$$

$$\leq \left\| \acute{\theta}_{\sim(2)}^{(n)} \right\|^2 + \alpha_n^2 l_n^2 \left\| A_n \right\|^2 ||Z_n||^2 + \alpha_n^2 l_n^2 \left| \left| A_n \right| \right|^2 ||H_{n,h}(\acute{\theta}_{\sim(2)}^{(n)})||^2 - 2\alpha_n < \acute{\theta}_{\sim(1)}^{(n)}\ , \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right) > +2 < \alpha_n \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n Z_n, \acute{\theta}_{\sim(2)}^{(n)} >$$

Using C_r-inequality $||a+b||^r \leq C_r(||a||^r + ||b||^r)$,

$$C_r = \begin{cases} 1 & , if\ r \leq 1 ; \\ 2^{r-1} & , if\ r > 1 \end{cases}$$

We will have

$\left|\left|\acute{\theta}_{\sim(2)}^{(n-1)}\right|\right|^2 \leq \left|\left|\acute{\theta}_{\sim(2)}^{(n)}\right|\right|^2 + \alpha_n^2 l_n^2 \left|\left| A_n \right|\right|^2 \left|\left| Z_n \right|\right|^2 2\alpha_n^2 l_n^2 \tau_n^2 ||A_n||^2 (||\theta_{(2)}^{(n)}||^2 + 1) - 2\alpha_n < \theta_{(2)}^{(n)}, \psi(\theta_{(1)}^{(n)}) A_n H_{n,h}(\theta_{(2)}^{(n)}) > + W_n$, ...(3.10)

With $E_{F_n} W_n = 0$, since $\alpha_n, A_n, \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right)$ and $H_{n,h}$ are F_n - measurable functions, then taking the conditional expectation of both sides of (3.10) with respect to F_n, we obtain $E_{F_n}||\acute{\theta}_{\sim(2)}^{(n+1)}||^2 \leq ||\acute{\theta}_{\sim(2)}^{(n)}||^2(1 + 2\alpha_n^2 l_n^2 \tau_n^2 ||A_n||^2) + 2\alpha_n^2 l_n^2 ||A_n||^2 \times (\tau_n^2 + 1/2 E_{F_n}||Z_n||^2) - 2\alpha_n < \acute{\theta}_{\sim(2)}^{(n)}, \psi(\acute{\theta}_{\sim(1)}^{(n)}) A_n H_{n,h}(\acute{\theta}_{\sim(2)}^{(n)}) >$ ....(3.11)

Therefore, by applying (nonnegative almost sure supermartingales convergence theorem) [11] , with $F_n = F_n$, $Z_n = ||\acute{\theta}_{\sim(2)}^{(n)}||^2$, $\beta_n = 2\alpha_n^2 l_n^2 \tau_n^2 ||A_n||^2$, $£_n = 2\alpha_n < \acute{\theta}_{\sim(2)}^{(n)}, \psi(\acute{\theta}_{\sim(1)}^{(n)}) A_n H_{n,h}(\acute{\theta}_{\sim(2)}^{(n)}) >$, and $£_n = 2\alpha_n^2 l_n^2 ||A_n||^2 (\tau_n^2 + 1/2 E_{F_n}||Z_n||^2)$, under (3.1),(3.7),and (3.8),satisfy the conditions of (nonnegative almost sure supermartingales convergence theorem)[11], we conclude that $\left\|\acute{\theta}_{\sim(2)}^{(n)}\right\|$ converges and is finite and

$$\sum_{n=1}^{\infty} \alpha_n (< \acute{\theta}_{\sim(2)}^{(n)}, \psi(\acute{\theta}_{\sim(1)}^{(n)}) A_n H_{n,h}(\acute{\theta}_{\sim(2)}^{(n)}) >) < \infty \quad ....(3.12)$$

Now (3.12) implies, under (3.3), that

$$\sum_{n=1}^{\infty} \alpha_n v_n l_n \left|\left|\acute{\theta}_{\sim(2)}^{(n)}\right|\right| \left|\left|H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right|\right| < \infty \quad ....(3.13)$$

But (3.13) and (3.4) imply that

$$\lim_{n\to\infty} inf \left|\left|\acute{\theta}_{\sim(2)}^{(n)}\right|\right| \left|\left|H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right|\right| = 0 \quad ....(3.14)$$

It remains to establish $\acute{\theta}_{\sim(2)}^{(n)} \to 0$, we do so by a contradiction. Suppose that here exists $£ \neq 0$ such that $P(\lim_{n\to\infty\infty} \left|\left|\acute{\theta}_{\sim(2)}^{(n)}\right|\right| = £) > 0$. Then choose $o < \Delta_1 < \Delta_2 < \infty$ such that $P(\Delta_1 < £ < \Delta_2) > 0$. Then for each, win the set $[\left|\left|\acute{\theta}_{\sim(2)}^{(n)}\right|\right| \to £$ with $\Delta_1 < £(w) < \Delta_2$, we have ,

$$\Delta_1 \leq \left|\left|\acute{\theta}_{\sim(2)}^{(n)}(w)\right|\right| \leq \Delta_2 \quad (3.15)$$

For n sufficiently large so that $(\Delta_1 \leq \left|\left|\acute{\theta}_{\sim(2)}^{(n)}\right|\right| \leq \Delta_2) > 0$. Thus, from (3.15),(3.11), and , (3.2) for n sufficiently large , we have

$$E_{F_n}||\acute{\theta}_{\sim(2)}^{(n+1)}||^2 \leq ||\acute{\theta}_{\sim(2)}^{(n)}||^2(1 + 2\alpha_n^2 l_n^2 \tau_n^2 ||A_n||^2) + 2\alpha_n^2 l_n^2 ||A_n||^2 \times (\tau_n^2 + 1/2 E_{F_n}||Z_n||^2) - 2\alpha_n l_n \Delta_1 v_n \rho_n \quad ....(3.16)$$

Apply (nonnegative almost sure supermartingales convergence theorem) [11]once more to (3.16) we conclude that $\sum_{n=1}^{\infty} \alpha_n v_n l_n < \infty$ which contradicts (3.4). Thus $\acute{\theta}_{\sim(2)}^{(n)} \to o \ as\ n \to \infty$.■

4- Asymptotic Normality of the Most General Stochastic Approximation Procedures

We prove a general theorem on the asymptotic normality of the most general stochastic approximation procedure in (2.1), based on FabianTheorem [7]

4.1-(<u>Asymptotic Normality Theorem)</u>

Let α, l_1, l_2, β be positive numbers and α, μ, γ be nonnegative numbers, Let Assumption (2.1) hold with

$= p - q$, $\alpha_n = a/n^{\alpha}$,

$0 < \alpha \leq 1$; $\beta = 1$ if $\alpha = 1$, $\beta = 0$ if $\alpha \neq 1$, ...(4.1)

The symmetric nonsingular $(P - q \times P - q)$ matrices A_n in $R^{(p-q\times p-q)}$ converge to a positive definite symmetric matrix A in $R^{(p-q\times p-q)}$ and the eigenvalues of $A_n(w)$ lie in$[v_n, \mu_n]$ for all w in **Ω**, where $v_n = (log(n+1))^{-1}, \mu_n = n^{\mu}$, $0 < \mu < \gamma$, $0 < \gamma < 1/2$. Moreover, let assumptions (2.2), (2.3) and (2.4) hold, the function h be continuous a.e, (F) and ∪ be differentiable at 0 in $R^{(p-q)}$ with

$$[DU(t)]_{t=0} = \left[\frac{\partial U^{(j)}(t)}{\partial t^{(i)}}\right]_{t=0} = K(h,f) \quad ....(4.2)$$

Where K(h,f) is a (p-qxp-q) positive definite symmetric matrix;

$(k(h,f))^{(ij)} =$

$\lim_{t^{(i)} \to o} (t^{(i)})^{-1} \int_{R^{p-q}} h^{(j)}(v) f(t-v) dv$ ....(4.3)

Set for $t \in R^{p-q}$

$\sum(t) = [\int_{R^{p-q}} [h^{(j)}(t+v) - u^{(j)}(t)][h^{(i)}(t+v) - u^{(i)}(t)] f(v) dv]_{i,j=1}^{p-q}$ ....(4.4)

Which is a (p-q x p-q) symmetric matrix with $\sum(0) = \sum_0(h,f)$ a diagonal matrix with diagonal elements

$\sigma_j^2(h.f)$; $j = 1,2, \ldots, p-q$ ....(4.5)

Assume that $\sum(t)$ is bounded, and, is continuous at

t=0 .... (4.6)

Let $l_1 \le \psi(\acute{\theta}_{\sim(1)}^{(n)}) \le l_2$ .... (4.7)

In addition; let

$F_n = \sigma(\acute{\theta}_{\sim(2)}^{(1)}, x_1, \ldots x_n ; h_1, \ldots h_{n-1} ; A_1, \ldots A_n)$, and for n=1,2,…..; M_n be $R^{(p-q \times p-q)} F_n$ - measurable satisfying

$\psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right) = M_n\left(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}\right)$ ....(4.8)

Assume that there exists M in $R^{(p-q \times p-q)}$ such that

$M_n \to M$ ....(4.9)

And AK(h,f)M is positive definite (4.10)

Set $\sigma_{r,n}^2 = E[\,||Z_n||^2 \times \{|||Z_n||^2 \ge rn^{\alpha}\}]$, $for\ r > 0$ ....(4.11)

Where $Z_n = -(h\left(\acute{\theta}_{\sim(2)}^{(n)}\right) - H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right))$, and Z_n is F_n-measurable random vector is R^{p-q} , and $h\left(\acute{\theta}_{\sim(2)}^{(n)}\right) = h_n$, Furthermore, assume that for all $r > 0$

$$\lim_{n\to\infty} \sigma_{r,n}^2 = 0, or\ \alpha$$

$$= 1\ and\ \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} \sigma_{r,j}^2 = 0$$

..................................(4.12)

Since AK(h,f)M is a symmetric (p-q x p-q) matrix , then there exists an orthogonal matrix $p \in R^{(p-q \times p-q)}$ such that p AK(h,f)M p = Λ is a diagonal matrix with the eigenvalues of AK(h,f)M lie on the main diagonal of Λ [8] . Assume that

$a\lambda > \frac{\beta}{2}$ $if\ \ \alpha = 1$..............................(4.13)

Where $\lambda = \min_{(i)}(\Lambda^{(ii)})$. Then $\acute{\theta}_{\sim(2)}^{(n)} \to \theta_2\ as\ n \to \infty$, and the asymptotic distribution of $n^{\alpha/2}(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)})$ is normal with mean=0 and

Covariance matrix = $a^2 l_2^2 p \sum p$, where $\sum^{(ij)} = (P\ A \sum_o (h,f) Ap)^{(ij)} (a\Lambda^{(ii)} + a\Lambda^{(jj)} - \beta)^{-1}$.

Proof:

Without loss of generally, set $\acute{\theta}_{\sim(2)} = 0$, we have

$H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right) = \cup\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right) =$

$(u^{(1)}\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right), \ldots, u^{(p-q)}\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right))$

and since $\cup$ is differentiable at o , then , [7], we have for $t \in R^{p-q}$,

$\cup(t) = D\cup(0)t + \eta(t)||t|| = K(h.f)t + \eta(t)||t||$ With $||\eta(t)|| \to o\ as\ t \to o,$

(Note DU(0) is a (p-q x p-q) matrix) . Therefore

$U\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right)$

$= K(h,f) H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)$

$+ \eta\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right) | \left|H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right| |$

With $\left|\left|\eta\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right)\right|\right| \to$

$0\ \ as\ H_n(\acute{\theta}_{\sim(2)}^{(n)}) \to 0$, Define the following (p-q x p-q) matrix-valued measurable function;

$$K_n(h,f) = \begin{cases} K(h,f) & if\ H_n\left(\theta_{(2)}^{(n)}\right) = 0 ; \\ K(h,f) + \eta\left(H_n(\theta_{(2)}^{(n)})\right) \dfrac{H_n(\theta_{(2)}^{(n)})}{\left|\left|H_n\left(\theta_{(2)}^{(n)}\right)\right|\right|} & if\ \ H_n(\theta_{(2)}^{(n)}) \ne 0 \end{cases}$$

And thus $H_{n,h}\left(\acute{\theta}_{\sim(2)}^{(n)}\right) = U\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right) = K_n(h.f) H_n(\acute{\theta}_{\sim(2)}^{(n)})$ ....(4.14)

Hence, using eq.(4.14) and under the given conditions, it can be easily checked that conditions of theorem (3.1) are satisfied and conclude that $\acute{\theta}_{\sim(2)}^{(n)} \to 0$. Thus, under (4.8), we get

$H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right) \to 0\ as\ \acute{\theta}_{\sim(2)}^{(n)} \to 0,$ and,

$K_n(h,f) \to K(h,f)\ \ (as\ \theta_{(2)}^{(n)} \to o)$ ....(4.15)

Now using eq. (4.14), we have $\acute{\theta}_{\sim(2)}^{(n+1)} = \acute{\theta}_{\sim(2)}^{(n)} - n^{-\alpha} a\psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n h_n$,

$n = 1,2, \ldots$

$= \acute{\theta}_{\sim(2)}^{(n)} - n^{-\alpha} a\psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n K_n(h,f) H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right) + n^{-\alpha} a\psi(\acute{\theta}_{\sim(1)}^{(n)}) A_n Z_n$ ....(4.16)

With $Z_n = -\left(h_n - U\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right)\right)$ $and \quad E_{F_n} Z_n = 0$

Let I be the (p-q x p-q) identity matrix, then eq. (4.16) by using eq. (4.8) , reduces to

$$\acute{\theta}_{\sim(2)}^{(n+1)} = \acute{\theta}_{\sim(2)}^{(n)} - n^{-\alpha} a A_n K_n(h,f) M_n \acute{\theta}_{\sim(2)}^{(n)} + n^{-\alpha} a\psi(\acute{\theta}_{\sim(1)}^{(n)}) A_n Z_n$$

$$= (I - n^{-\alpha} a A_n k_n(h,k) M_n)\acute{\theta}_{\sim(2)}^{(n)} + n^{-\alpha} a\psi(\acute{\theta}_{\sim(1)}^{(n)}) A_n Z_n$$

Then applying (Fabian theorem)[7] ,we have satisfied the conditions, with

$U_n = \acute{\theta}_{\sim(2)}^{(n)}$, $\quad \beta = \alpha$

$\Gamma_n = a\, A_n K_n(h,f) M_n \to a\, AK(h,f)M = \Gamma$

$T_n = 0 = T$, $V_n = Z_n$, $\Sigma = \Sigma_o(h,f)$

$\phi_n = a\psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n \leq a l_2 A_n \to a l_2 A = \phi$

To satisfy the remaining condition of(Fabian theorem)[7],we have

$E_{F_n} V_n V_n = E_{F_n} Z_n Z_n = \Sigma(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right))$

$\lim_{n\to\infty} E_{F_n} V_n V_n = \lim_{n\to\infty} \Sigma\left(H_n\left(\acute{\theta}_{\sim(2)}^{(n)}\right)\right) = \Sigma_o(h,f)$(4.17)

And for $t \in R^{(p-q)}$,

$E\left(t\left(E_{F_n} Z_n Z_n - \sum_o(h,f)\right)t\right) = t\left(E_{F_n} Z_n Z_n - \sum_o(h,f)\right)t \to 0$(4.18)

Define for $r > 0$,

$\rho_{r,n}^2 = E[||V_n||^2 \times \{V_n||^2 \geq rn^{\alpha}\}]$

$= E[||Z_n||^2 \times \{||Z_n||^2 \geq rn^{\alpha}\}]$(4.19)

By eq.(4.12),we have

$\rho_{r,n}^2 \to 0 \; as \; n \to \infty, \; or \; \alpha = 1 \; and \; n^{-1}\sum_{j=1}^{n} \rho_{r,j}^2 \to 0$(4.20)

Then we conclude that the asymptotic distribution of

$n^{\alpha}/_2(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)})$ Is normal with

$Mean = (aAK(h,f)M - \left(\frac{\beta}{2}\right) I)^{-1} 0 = 0,$

And covariance matrix = $a^2 l_2^2 P \sum P$,

Where $\sum^{(ij)} = (PA\sum_o(h,f)AP)^{(ij)}(a\Lambda^{(ii)} + a\Lambda^{(jj)} - \beta)^{-1}$ ■

5- Confidence intervals For the Vector of Nonlinear Parameters

To construct confidence intervals for the vector of nonlinear parameters, $\acute{\theta}_{\sim(2)}$, of the most generalized model (1.3) . Using the result of theorem (4.1),we have

$n^{\alpha/2}\left(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}\right) \xrightarrow{L} N(0, a^2 l_2^2 P \sum P) \; as \; n \to \infty$,

Where $\acute{\theta}_{\sim(2)}^{(n)}$ is the nth iterative of the most general stochastic approximation procedure given by (1.4). Then the joint $100(1-\alpha)\%$ confidence intervals for the vector of nonlinear parameters, $\acute{\theta}_{\sim(2)}$, is given by: $\frac{n^{\alpha/2}\left(\theta_{(2)}^{(n)} - \theta_{(2)}\right)(\theta_{(2)}^{(n)} - \theta_{(2)})}{a\, l_2 \left|\sqrt{P\sum P}\right|} \leq Z_{1-\alpha}$, i.e.,

$n^{\alpha/2}\left(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}\right)\left(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}\right) \leq a\, l_2 Z_{1-\alpha} ||\sqrt{P\sum P}||$,

Where $\sum^{(ij)} = (P\, A \sum_o(h,f)AP)^{ij}(a\Lambda^{(ii)} + a\Lambda^{(jj)} - \beta)^{-1}$,

And $Z_{1-\alpha}$ is the $1-\alpha$ point ("upper α-point") of the normal distribution.

6- Conclusion

We consider the most general nonlinear regression model

$Y(x) = g(\theta; x) + \varepsilon$, when the regression function, as far as the parameters are concerned, is composed of linear and nonlinear regression function, i.e., $g(\theta; x) = \psi(\theta_{(1)}) g_1(\theta_{(2)}; x)$. We study the most general stochastic approximation procedure to estimate $\acute{\theta}_{\sim(2)}$, given by: $\acute{\theta}_{\sim(2)}^{(n+1)} = \acute{\theta}_{\sim(2)}^{(n)} - \alpha_n \psi\left(\acute{\theta}_{\sim(1)}^{(n)}\right) A_n h_n$,

$n = 1,2, \ldots$

We prove that almost sure convergence of $\acute{\theta}_{\sim(2)}^{(n)}$, i.e., $\acute{\theta}_{\sim(2)}^{(n)} \to \acute{\theta}_{\sim(2)}$, $as \; n \to \infty$ and the asymptotic normality of $n^{\alpha/2}\left(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}\right)$, for $o < \alpha \leq 1$, which is given as follows $n^{\alpha/2}(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)})$ is normal with mean=0, and covariance matrix= $a^2 l_2^2 P \sum P$, where $\sum^{(ij)} = (P\ A \sum_o(h,f)AP)^{(ij)}(a\Lambda^{(ii)} + a\Lambda^{(jj)} - \beta)^{-1}$

Finally, we construct general confidence intervals for the vector of nonlinear parameters, $\acute{\theta}_{\sim(2)}$, which is given by:

$$n^{\alpha/2}\left(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}\right)\left(\acute{\theta}_{\sim(2)}^{(n)} - \acute{\theta}_{\sim(2)}\right) \leq a\ l_2 Z_{1-\alpha}\left|\left|\sqrt{P\sum P}\right|\right|$$

References

[1] Sami, A,N,."Transformation of Observation in Stochastic Approximation", Ann. Statist. **1,6, pp 1158-1174, 1999.**

[2] Sami, A, N, "On optimal multidimensional stochastic approximation", A unified approach 1 ,The proceeding of math, and computer science conference, UNESCO and Alexandria Univ., Nov**. 2002, pp 393-413, 2002.**

[3] Sami, A, N, "On optimal multidimensional stochastic approximation procedures", Florida State university statistical Report no. M-603, October, submitted to Communications in Statistics 2007.

[4] Sami, A, N," On estimation Via optimal stochastic approximation procedures", Statistical Theory and Data Analysis, K. Maturity (Editor) Elsevier Science Publishers B.V. (North Holland), 2009.

[5] Bartle,R.G. "The Elements of real Analysis, John Wiley & Sons, Inc., New York, 2001.

[6] Draper, N.R. and Smith, h**,** "Applied Regression Analysis", john Wiley & Sons,Inc.,New York.2004

[7] Fabian,V**.**"On asymptotic normality in stochastic approximation", Ann, Math. Statist. 39**, pp 1327-1332, 2000.**

[8] Graybill, F.A**."** Introduction to Matrices with approximations in Statistics", Wadsworth publishing Co., Inc., California, 1996.

[9] Lawton, W. H. and Sylvestre, E.A. "Elimination of linear parameters in nonlinear regression", Teahnometrics, 13, p 461-467, Discussion**, pp 477-481, 1998.**

[10] Mc Cullagh, p. and Neider, J, K**.** "Generalized Linear Models", Chapman-Hall-London, 2008.

[11] Robbins, H. and siegmund, D, "a convergence theorem for nonnegative almost supermartingales and some approximations", Proc. Symp. Optimizing Methods in Statistics (J.S.Rustagi,ed.). p**p 233-257**, Academic Press, New York, 2005.

الخلاصة

في هذا البحث تم دراسة نموذج الانحدار الغير خطي العام $Y(x) = \psi(\theta_{(1)})g_1(\theta_{(2)};x) + \varepsilon$ باستخدام طريقة جديدة عامة للتقريب العشوائي الامثل. برهنا ان التقديرات للمتجه $\theta_{(2)}$ لها نهايات تقريبية أكيدة ومحاذيات طبيعية. قمنا ببناء طريقة لإيجاد فترات الثقة العامة لمتجه الوسائط الغير خطية كما تم تقديم نموذج عام للانحدار غير الخطي وتم استخلاص بعض المميزات العامة للنموذج المقترح.