The primary intent of this publication, Technical Report, is to share original work as quickly as possible. Therefore, articles which appear are not reviewed as is the usual practice with most journals. The authors alone are responsible for the content, interpretation of data, and opinions expressed in the articles. All communications concerning the articles should be addressed to the author c/o:
COMPUTING ESTIMATES BY THE NEWTON-RAPHSON TECHNIQUE

Jimbo Henri Claver¹ and Takeru Suzuki²

¹ Dept of prediction and Control, Institute of Statistical Mathematics, 4-6-7 Minami Azabu, Minato-ku, Tokyo 106-8569, Japan. E-mail: jimboc@ism.ac.jp
RISE member

² Dept of Mathematical Sciences, Waseda University, Tokyo-Japan
RISE-member

November 14, 2003

ABSTRACT

We investigate an estimating equation matrix of derivatives of a function \( w \). We define the estimating function \( w_n \) as the function used for that purpose. The Newton-Raphson procedure for the estimates and the Lusternik method for increasing the convergence of the iteration process are discussed in this paper.

Keywords: Newton-Raphson Algorithm, Lusternik Method, M, AM, GM, AGM Estimates Estimating Equation, Quadratically Convergent Process.

1. INTRODUCTION

The construction of estimates has always been a central problem in parametric and semiparametric statistic and model identification. For many decades different authors at different times have brought valuable contributions to this topics; we have for example Fisher (1956), Bickel (1985), Bickel et al (1993), Le Cam (1972), Heckman (1979), Hardy, Litterwood and Polya (1952), Hasmiski and Ibragimov (1983), Wilsk (1962) and others. A special attention has been focused on the paper of Wilson, which for the first time has introduced iterative procedure in estimation problems. It is well known that the Newton Raphson provides higher convergence properties for quadratically convergent processes. Conceptually there are essentially only two basic approach to the construction of estimates of Euclidian parameters: The first is by finding a smooth extension of the parameter and the second is by finding a good estimate of the parameter; even in the case of abstract parameters the same approach is applied. This is why in the next part of this work we would like to investigate essentially the iterative characteristic of estimates of Euclidean parameters. This
paper is organized as follow; in section one we give some definitions of M, GM and AGM estimates, in section two we show how to construct good estimating equations \( w_n \) for Euclidean parameters and pose (or) solve the non linear optimization problem, in section three we show our computer program for the Newton Raphson procedure for the estimate with simulation using the MATLAB and the TOPIX data, in section four we use the Lusternik method to improve the rate of convergence of the iteration process and lastly in section five we end with some conclusive remarks.

1.1 M; AM; GM; AGM ESTIMATES

1.1.1 DEFINITIONS

1.1.2 M; AM; ESTIMATES

In this section the familiar M-estimate is described and will serve as convenient introduction to the Generalized M-estimates (Cf. Bickel et al. (1993)). Let \( P = \{\text{all probabilities on } \mathbb{R} \} \) dominated by Lebegue measure; \( M_0 \supset \{\text{probability distributions with finite support}\} \) and \( \nu(P) \) is defined like \( \nu : P \to \mathbb{R}^m \). Suppose that \( \phi : \mathbb{R}^m \to \mathbb{R}^m \) and let \( w(\nu, Q) = \int \phi(x, \nu) dQ(x) \) be defined for all \( Q \in M_0 \supset P \). We assume that for each \( p \in P \), our parameter of interest \( \nu(p) \in \mathbb{R}^m \) satisfies:

\[
(\nu(p), \nu(p)) = 0
\]

We call any root \( w(\nu, P_n) = n^{-1} \sum_{i=1}^{n} X_i \) an M-estimate where \( P_n \) means the empirical distribution function based on observations \( \{X_i\} \). An M-estimate is called an asymptotic M-estimate or AM-estimate if it satisfies \( w(\hat{\nu}, P_n) = o_p(n^{-1/2}) \).

1.1.3 GM; AGM ESTIMATES

Suppose \( M_0 \supset P \) and all distributions with finite support as in section 1.1.1. Suppose that \( w_n, w \) map \( \mathbb{R}^m \times M_0 \to \mathbb{R}^m \) and that:

(a) \( w_n(\nu, P) = w(\nu, P) + o(1) \), for all \( P \in M_0 \), all \( \nu \),

(b) \( w(\nu(P), P) = 0 \), for all \( P \in P \).

For the empirical distribution \( P_n \), we introduce the notation \( w_n(\nu) = w_n(\nu, P_n) \). Then, \( \hat{\nu}_n \) is a generalized M-estimate or a GM-estimate of \( \nu(P) \).

\[
\text{if } w_n(\hat{\nu}_n) = 0 \quad (1) \\
\text{if } w_n(\hat{\nu}_n) = o_p(n^{-1/2}) \quad (2)
\]
for all \( P \in P \), we say that \( \hat{\nu}_n \) is an asymptotic generalized M-estimate, or AGM-estimate of \( \nu \).

**REMARKS:**

Extension of these estimates to the infinite dimensional parameters is feasible. Estimates of this type were first considered by Fillipova (1962) under regularity conditions considerably more stringent than the ones we present in this paper. Some regularity conditions are necessary to say anything is evident since any estimate \( \hat{\nu}_n(P) \) can be considered as a GM-estimate by taking \( w_n(\nu, P) = \nu - \hat{\nu}_n(P) \). In this respect, the AGM-estimate is an extension of the M-estimate [see for example Huber (1964)].

**EXAMPLES**

- Hodges-Lehman Estimate
- Estimates for the Hasminski-Ibragimov model
- Estimates for the elliptic model
- Estimates for the semi-parametric model
- Other examples of partially efficient M-estimates are rare

**2. NONLINEAR OPTIMIZATION**

**2.1. ILL-POSED PROBLEM**

The generalized M-estimate formulation leads naturally to condition for linearization of \( w_n(\nu, \cdot) \), consistency, asymptotic linearity, and normality for \( \nu(P) \). Consistent solution for (1) should be \( \sqrt{n} \) consistent, asymptotically linear, normal etc… this justifies why the above nonlinear equation have to be solved iteratively. However the problem connected to (1) is an ill posed problem in the sense that it may have infinitely many number of solutions when this exist, and the best solution is obtained only if we start enough close to the unknown \( \nu(P) \) that is if we are lucky or have available crude consistent starting value.

**2.1.1 NEWTON RAPHSON PROCEDURE FOR SOLVING THE ESTIMATING EQUATION**

Recall the estimating equation:

\[ w_n(\nu_n) = 0 \]  

The solution(s) \( \nu_n = \hat{\nu}_n, (\hat{\nu}_n = \nu_n^\infty) \) will be found by a constructive iterative approach combined with the one step ahead method, but let us first remind briefly the one step ahead method (OSAM). The above method says that if we have consistent estimate \( \tilde{\nu}_n \) of \( \nu_n \), we can construct a consistent AGM-estimate. In those conditions we have \( \hat{w}^* = \hat{w}(P) + o_P(1) \),
\( (\hat{w}(p) = \hat{w}(v, p)) \) which is a consistent estimate of \( \hat{w} \) and the one step ahead estimate in that case is:

\[
\hat{w}_n = \hat{w}_n - (\hat{w}^*)^{-1}w_n(\hat{w}_n)
\]

Where \((\hat{w}^*)^{-1}\) is the general inverse of \(\hat{w}^*\) and \(\hat{w}_n\) the first Newton-Raphson iteration for solving (1) starting from \(\hat{w}_n\). We claim that \(\hat{w}_n\) to be consistent. It is then possible to take \(\hat{w}^* = \hat{w}_n(\hat{w}_n)\) and then iterate the process as:

\[
\begin{align*}
T_n^{(0)} &= \hat{w}_n \\
T_n^{(j+1)} &= T_n^{(j)} - w_n(T_n^{(j)})\hat{w}_n^{-1}(T_n^{(j)}) & ; \quad j = 0, 1, 2, 3, \ldots
\end{align*}
\]

We iterate to the limit \(j = \infty\), which will give us the GM-estimate. But the limit may not exist and if it does, may not satisfy the consistency condition. It is however under an additional uniform convergence condition on \(\hat{w}_n(v)\) to establish the asymptotic linearity of \(T_n^{(\infty)}\) with probability tending to one. In mathematical interpretation we have the following:

\[
\Pr\{T_n^{(\infty)} \text{ exists and equals } \nu_n^{\infty}\} \to 1; \quad n \to \infty \quad \text{(Cf. Bickel \textit{et al} (1993))}
\]

\section*{3. COMputation with MATLAB}

The MATLAB software is used for our computation and an application to the TOPIX data is also carried out. The results of our computation and the program can be seen at the end of this paper.

\section*{4. LUSTRENIK APPROACH}

As it is already well known one of the major problem of the Newton-Raphson algorithm is the convergence of the iteration process. At some steps, it may happen that the iteration process diverges. This can be due to the starting value which is wrongly choose [not enough close to the real (exact) value or the speed or rate of convergence sufficiently low. However to overcome the second obstacle we need to increase the rate of convergence and this can be done by applying the Lusternik Method.

\subsection*{4.1 Application}

For \(m\) sufficiently large, we put \(T_n = T_n^{(m)}\) in (4) and evaluate the error \(T_n - T_n^{(m)}\) that is
used for our iteration. Under the following convergence condition of the new iteration process, which is: for any arbitrary function \( f \); \[ f(w_n, T_n) \] < \( \varepsilon \), we read \( T_n \) as

\[
T_n = \lim_{m \to \infty} T_n^{(m)} = T_n^{(0)} + \sum_{k=1}^{m} (T_n^{(k)} - T_n^{(k-1)}).
\]

On the other hand \( T_n^{(m)} = T_n^{(0)} + \sum_{k=1}^{m} (T_n^{(k)} - T_n^{(k-1)}) \)

and

\[
T_n - T_n^{(m)} = \sum_{m+1}^{\infty} (T_n^{(k)} - T_n^{(k-1)}) = \left[ T_n^{(m+1)} - T_n^{(m)} \right] + \left[ T_n^{(m+2)} - T_n^{(m+1)} \right] + \ldots.
\]

By continuing this approximation based on the successive difference, we obtain at the end a supplement term \( \beta \) such that:

\[
\beta = \frac{T_n^{(K+1)} - T_n^{(K)}}{1 - \gamma}
\]

where

\[
\gamma \approx \frac{(T_n^{(m)} - T_n^{(m-1)})}{(T_n^{(m-1)} - T_n^{(m-2)})}, \quad i = 1, 2, 3, \ldots, n
\]

which improves the convergence of the iteration procedure. However by using the method of inner product we obtain a more precise value of \( \gamma \); we then have

\[
\gamma = \frac{(T_n^{(m)} - T_n^{(m-1)}, T_n^{(m)} - T_n^{(m-1)})}{(T_n^{(m-1)} - T_n^{(m-2)}, T_n^{(m-1)} - T_n^{(m-2)})}
\]

**COMPUTER PROGRAM**

%Newton-Raphson Algorithm and application to Topix Data
%By Jimbo 2003

%ARMA (p,q)MODEL (jim4)

clear all;
close all;
load Topix.dat;  %Data -ascii;
p=3;
q=5;
\begin{verbatim}

z=Topix(1:end,2);
N=size(z,1);
l=mean(z)
x=z-mean(z)

% x=x';
% pause

Npq=N-(p+q);
y1=ones(Npq,1)

%% A=zeros(N-p,p);
% s=0
% for e=1:(p+q);
%   D(e,:)=x(e:e+p+q-1);
%   D(:,e)=x(p+q+1-e:N-e);
%   pause
% end
Y=x(p+q+1:N)
fi=pinv(D'*D)*(D'*Y)
hatY=D*fi
E=Y-hatY;
figure
subplot(2,1,1),plot([1:N-(p+q)],Y,[1:N-(p+q)],hatY,'r:') %plot data & estimate
subplot(2,1,2),plot(E); %plot prediction error
figure;
hist(E,50); % histogram of pred. error
figure;
hist(x,50); % histogram of data
E2=E'*E;
Var=E2/(N-(p+q));

h=100;
PAI=3.1415;

s=0
% f=[0.5:-0.5]
for k=1:p+q
  pa=3.1415;
delta=0.001;
h=0:500;
f=h*delta;
  sk = fi(k)*exp(-2*pa*i*k*f)
  s=s+sk
% pause
\end{verbatim}
end
dem=abs(1-s).*abs(1-s);
w=2*Var./dem;                           % spectrum
semilogy(w)
figure;
plot(w);

%Estimate the autoregressive parameters hata(phi)
%Estimate the moving average parameters hat(teta)

K=p+q
for a=1:K
   u=0
   for b=1:N-a
      u=var(x)*x(b+a)+u;
   end
   acf(a)=(1/N)*u
end

newacf=[var(x),acf]                  % new autocovariance function
figure
plot(acf)
   ro=(acf)/var(x)                      % autocorrelation function
   % for k=1:p+1
figure
subplot(2,1,1),plot(acf)
subplot(2,1,2),plot(ro)
figure
plot(newacf)
M=toeplitz(ro(q:p+q-1))               % Estimate of the autoregressive parameters
Mat=triu(M)
S=toeplitz(ro(q:-1:abs(q-p+1)))
Sat=tril(S)

Vap=M
G=pinv(Vap)
phi=G*(ro(q+1:p+q))'                 % Yulke-Walker equation
% end
figure
plot(phi)
modacf=0

newfi=[-1;fi]
for k=1:p+1
    for r=1:p+1
        for m=1:p+1
            % Estimate of the moving average parameters hatTeta
            ind=abs(e+k-f)
            if r+k-m==0
                dr=newfi(r).*newfi(m)*newacf(2);
            else
                dr=newfi(r).*newfi(m)*newacf(abs(r+k-m));
            end
            modacf=modacf+dr
            % "moacf modified autocovariance"
            end
            end
            if p==0
                newfi=acf(k)
            end
            l=0
            if r-m==0
                l=l+newfi(r)*newfi(m)*newacf(2)
            else
                l=l+newfi(r)*newfi(m)*newacf(abs(r-m))
            end
        end
    end
end

% Newton Raphson begins here

% TAU=[];
% wax=zeros(1,q);
tau1(1)=(l)^1/2
tau=[tau1(1),wax]
F=ones(1,q+1)

% while abs(F)<eps

ifar=toeplitz(tau);
wav=triu(ifar);
amb=rot90(rot90(tau));
tor=toeplitz(amb);
hit=triu(tor);
wac=rot90(hit);
T=wac+wav;
% Pause
for m=1:q+1
    fsmall=0;
    for r=1:q-m+1
        fsmall=fsmall+tau(r)*tau(r+m);  % general iteration
    end

    fsmall=fsmall-newacf(m);
    F(m)=fsmall;
end

% F,T->H
H=pinv(T)*F';

%TAU=[TAU;tau] ;
while abs(F(m))<eps
    tau=tau-H';
    F
    pause
    T
    pause  %make a new tau
end

hatTeta=-(TAU)*pinv(tau)  %estimate hatTeta as ratio -TAU/tau

daad=0

for m=1:q+1
    if p>0
        daad=daad+newacf(m)  %estimate of the overall constant hatTeta0
    else daad=newacf
end
faar=I*(1-daad)

navar=0
for m=1:q+1
    if q==0
        navar=navar+newacf(m)*acf(m)
    else  q>0
        navar=(tau).^2  %estimate of the white noise variance
    end
end
5. CONCLUSION

The Newton Raphson approach is efficient only when the starting value of the preliminary estimate $T_n^{(0)}$ is given sufficiently close to the exact consistent estimate $\hat{\nu}_n$. But the exact consistent estimate is unknown. In this note we have proposed an approach, which is globally able to give a starting value corresponding to each estimate, and show how to increase the convergence of the iterations process by the Newton-Raphson algorithm. We have also successfully applied the Newton Raphson Technique to real data (Topix Data) and the obtained results are reasonable.

Comments

The two main innovations in this work: The use of the Lusternisk technique to increase the speed of convergence of the iteration process and the application of the Newton-Raphson algorithm to real data; the efficiency of our result can be seen in the figures at the end of this paper.

Acknowledgments.

H.C. Jimbo is gratefully acknowledges the support by the Japan Society for the Promotion of Science (JSPS), the Institute of Statistical Mathematics and the RISE-Waseda University.

REFERENCES


Appeared as a Technical Report on the home page of the Advanced Research Institute of Science and Engineering, Waseda University, on December 1, 2003.