Sankhyā : The Indian Journal of Statistics
1993, Volume 55, Series A, Pt. 2, pp. 312–324.

ON ASYMPTOTIC DISTRIBUTION OF GENERALIZED M-ESTIMATES

By J. K. GHOSH and S. N. JOSHI
Indian Statistical Institute

SUMMARY. Let $\{X_n(t)\}$ be a sequence of stochastic processes with continuous paths. In this note, the asymptotic distribution of a suitably chosen root of $X_n(t) = 0$ is investigated. This leads to asymptotic distribution of M-estimates.

1. Introduction

Let F_n be the empirical distribution function (d.f.) corresponding to a sample of size n from a d.f. F. Let t_0 be a solution of $\lambda_F(t) = 0$ where

$$\lambda_F(t) = \int \psi(x, t)dF(x).$$

An M-estimate T_n (of t_0) corresponding to the kernel function ψ is a solution of the equation

$$\lambda_{F_n}(t) = 0. \quad \ldots \quad (1.1)$$

Under certain conditions on the function F, the asymptotic distribution of T_n can be obtained (See Boos and Serfling (1980) and Serfling (1980) Chapter 7). Instead of equation (1.1) one may have a more general form of equation

$$X_n(t) = 0 \quad \ldots \quad (1.2)$$

and a solution $\hat{\theta}$ of (1.2) as a estimate of a unknown parameter θ. We call this estimate $\hat{\theta}$ a generalized M-estimate (of θ). The estimate $\hat{\tau}$ of Nguyen, Rogers and Walker (1984) is of this type. Basu, Ghosh and Joshi (1988) (henceforth abbreviated as BGJ) obtained the asymptotic distribution of $\hat{\tau}$ above. In doing so, BGJ implicitly developed a technique which has wider applicabilities. In this note we consider the equation (1.2) where $\{X_n(t)\}$ is a stochastic process with continuous sample path and show how one can study the asymptotic distribution of $\hat{\theta}$ by explicitly writing down the techniques of BGJ and developing them further.

A related problem of asymptotic behaviour of One-Step M-estimates can also be tackled using the above mentioned techniques.

Paper received. August 1990; revised February 1991.
AMS (1990) *subject classification*, Primary 62F12; secondary 60K10,
Key words. M-estimates, asymptotic distribution, stochastic processes.

2. Preliminaries and notations

Let $\{X_n(t)\}$ be a sequence of stochastic processes having continuous sample paths with $t \in (a, b)$ and let θ be the parameter to be estimated, $\theta \in (a\ b)$. Let $C(-\infty, \infty)$ be the space of continuous function on $(-\infty, \infty)$; let $C(-\infty, \infty)$ be endowed with the topology of uniform convergence on compacta (u.c.c.). For details regarding weak convergence, tightness etc. in $C(-\infty, \infty)$ see Sen (1981).

For $f \in C(-\infty, \infty)$, let

$$S_f = \{t : f(t) = 0\}$$

and define functions T_1 and T_2 as

$$\begin{aligned} T_1(f) &= \sup\{t : f(t) = 0\} && \text{if } S_f \neq \phi \text{ and } S_f \text{ is bounded above} \\ &= c && \text{otherwise } (c \text{ is a constant}) \\ T_2(t) &= \inf\{t : f(t) = 0\} && \text{if } S_f \neq \phi \text{ and } S_f \text{ is bounded below} \\ &= c && \text{otherwise} \end{aligned}$$

Define a stochastic process $W_n(.)$ on $C(-\infty, \infty)$, a process obtained by rescaling $n^{\beta} X_n(.)$ (for some $\beta > 0$) at θ, by

$$\begin{aligned} W_n(h) &= n^{\beta} X_n(\theta + n^{-\beta} h) && \text{if } |h| \leqslant \log n \\ &= n^{\beta} X_n(\theta + n^{-\beta} \log n) && \text{if } h \geqslant \log n \\ &= n^{\beta} X_n(\theta - n^{-\beta} \log n) && \text{if } h \leqslant -\log n. \end{aligned} \quad \ldots \quad (2.1)$$

Equation (1, 2) may have multiple solutions ; we use an estimate $\hat{\theta}_c$ of θ to overcome this difficulty.

Let

$$\delta_n = n^{-\beta} (\log n)^{\gamma} \text{ and } \gamma = 1/2$$

(The arguments in the proof of Theorem 2.1 can be modified so as to work with any $\gamma > 0$).

Let $\hat{\theta}_c$ be such that

$$P(|\hat{\theta}_c - \theta| < \delta_n) \to 1 \text{ as } n \to \infty. \quad \ldots \quad (2.2)$$

With $\epsilon_n = \frac{1}{2} n^{-\beta} (\log n)^{\frac{1}{2}}$ let

$$I_n = [\hat{\theta}_c - \epsilon_n,\ \hat{\theta}_c + \epsilon_n] \cap \{t : X_n(t) = 0\}.$$

Now we define our estimate $\hat{\theta}$ of θ ;

$$\begin{aligned} \hat{\theta} &= a \text{ point in } I_n \text{ closest to } \hat{\theta}_c \text{ if } I_n \neq \phi \\ &= \hat{\theta}_c \text{ otherwise.} \end{aligned}$$

Condition (2.2) for $\hat{\theta}_c$ may be too demanding ; at the end of Section 3 (see Lemma 3.1) we indicate how to obtain $\hat{\theta}_c$ satisfying (2.2) when a consistent estimate of θ is available.

3. Main results

Proof of Theorems 3.1, 3.2 and 3.3 was essentially contained in the treatment of the special case considered in BGJ, though without mentioning the results explicitly ; for the sake of completeness, below we give sketches of the proofs as well.

Theorem 3.1. *Let $W(.)$ be a stochastic process on $C(-\infty, \infty)$ such that*

(*i*) $W_n(.) \xrightarrow{w} W(.)$;

(*ii*) $W(h) = 0$ *has a unique solution, say, a r.v.* Y ;

(*iii*) T_1 *and* T_2 *are continuous w.p.* $1\,[W(.)]$ *and* $P(Y = c) = 0$.

Then $\quad n^{\beta}(\hat{\theta}-\theta) \xrightarrow{w} Y.$

Remark 1. In what follows we take $\beta = 1/2$; all the statements of the results can be modified by replacing $\sqrt{n}$ by n^{β} and can be proved analogously.

Proof. Let

$$A_n^{(1)} = \{|T_i(W_n(.))| \leqslant (\log n)^{1/3},\ i = 1,2\},$$

$$A_n^{(2)} = \{T_1(W_n(.)) \neq c\} \cap \{T_2(W_n(.)) \neq c\},$$

$$A_n^{(3)} = \{|\hat{\theta}_c - \theta| \leqslant \delta_n\}$$

and $$D_n = \bigcap_{i=1}^{3} A_n^{(i)}.$$

It is easy to see that

$$P(A_n^{(i)}) \to 1 \text{ as } n \to \infty \text{ for } i = 1, 2, 3$$

and hence

$$P(D_n) \to 1 \text{ as } n \to \infty, \qquad \dots \quad (3.1)$$

Note that on D_n, $W_n(h) = 0$ for at least one h between $-(\log n)^{1/3}$ and $(\log n)^{1/3}$ and hence by (2.1)

$$X_n(t) = 0 \text{ for at least one } t \text{ between}$$

$$\theta-(\log n)^{1/3} \text{ and } \theta+(\log n)^{1/3} \qquad \dots \quad (3.2)$$

Now note that there exists a n_0 such that for all $n \geqslant n_0$

$$\{|\hat{\theta}_c-\theta| < n^{-1/2}(\log n)^{1/2}\}$$
$$\Rightarrow \{\theta-n^{-1/2}\log n \leqslant \hat{\theta}_c-\epsilon_n \leqslant \theta-n^{-1/2}(\log n)^{1/3}$$

and

$$\theta+n^{-1/2}(\log n)^{1/3} \leqslant \hat{\theta}_c+\epsilon_n \leqslant \theta+n^{-1/2}\log n\} \quad \ldots \quad (3.3)$$

In view of (2.1) definition of $\hat{\theta}$, (3.1) and (3.3) we have for all $n \geqslant n_0$ on D_n,

$W_n(\sqrt{n}(\hat{\theta}-\theta)) = 0$ and the set $\{h : W_n(h) = 0\}$ is bounded.

Thus for all $n \geqslant n_0$ on D_n,

$$T_2(W_n(.)) \leqslant \sqrt{n}(\hat{\theta}-\theta) \leqslant T_1(W_n(.)). \quad \ldots \quad (3.4)$$

Note that continuity of T_1-T_2 (w.p.1 $[W(.)]$) follows from continuity of T_1 and T_2 ; using continuity of T_1-T_2 along with (i) and (ii) we have

$$T_1(W_n(.))-T_2(W_n(.)) \xrightarrow{p} 0. \quad \ldots \quad (3.5)$$

Proof of the theorem is now completed by using (3.1), (3.4). (3.5) and the fact $T_2(W_n(.)) \xrightarrow{w} Y$.

It may be noted that the rescaling technique used above is due to Prakasa Rao (1968, 1986).

Note that the conditions of Theorem 3.1 are in terms of $W_n(.)$; Theorem 3.2 below gives conditions on $X_n(.)$ which ensure condition (i) above.

Theorem 3.2. *Let for some* $\epsilon > 0$, *as a stochastic process on* $C[\theta-2\epsilon, \theta+2\epsilon]$,

$$\sqrt{n}(X_n(.)-X(.)) \xrightarrow{w} G(.)$$

where $G(.)$ *is a stochastic process on* $C[\theta-2\epsilon, \theta+2\epsilon]$ *and* $X(.)\epsilon\, C[\theta-2\epsilon, \theta+2\epsilon]$ *is such that*

(*i*) $X(t)$ *is non-random*

(ii) $X(\theta) = 0$

(iii) *right and left derivatives of* $X(t)$ *at* θ, $\dot{X}(\theta+)$ *and* $X(\theta-)$ *respectively, exist.*

Then hypothesis (i) *of Theorem* 3.1 *holds with*

$$W(h) = h\,\dot{X}(\theta-)+G(\theta) \quad \text{if } h \leqslant 0$$
$$= h\,\dot{X}(\theta+)+G(\theta) \quad \text{if } h \geqslant 0.$$

Proof. Let

$$Y_n(.) = n^{1/2}(X_n(.)-X(.))$$

and $W_{1n}(.)$ be the process obtained by rescaling $Y_n(.)$ at θ (see (2.1). Note that, using tightness of $Y_n(.)$, given $h, \eta > 0$ and $\epsilon > 0$ we can get n_0 such

$$P(|W_{1n}(h)-W_{1n}(0)| > \epsilon) \leqslant \eta \ \forall \ n \geqslant n_0$$

i.e. $W_{1n}(h)-W_{1n}(0) \xrightarrow{p} 0$.

Now note that

$$W_{1n}(0) = Y_n(\theta) \xrightarrow{w} G(\theta)$$

Thus $$W_{1n}(h) \xrightarrow{w} G(\theta) \text{ for every fixed } h.$$

Convergence of finite dimensional distributions of $W_{1n}(.)$ (as a process in $C[-j, j]$) can now easily be established ; its tightness follows easily by using tightness of $Y_n(.)$. Thus writing

$$G_1(.) = G(\theta) \text{ w.p. } 1$$

we have, as a process in $C(-\infty, \infty)$.

$$W_{1n}(.) \xrightarrow{w} G_1(.). \quad \ldots \quad (3.7)$$

Now note that

$$W_{1n}(h) = W_n(h)-X_{1n}(h)$$

where $W_n(h)$ is as in (2.1) and

$$\begin{aligned} X_{1n}(h) &= n^{1/2} X(\theta+n^{-1/2}h) && \text{if } |h| \leqslant \log n \\ &= n^{1/2} X(\theta+n^{-1/2} \log n) && \text{if } h \geqslant \log n \\ &= n^{1/2} X(\theta-n^{-1/2} \log n) && \text{if } h \geqslant -\log n. \end{aligned}$$

Let

$$\begin{aligned} X_1(h) &= h\dot{X}(\theta+) \quad \text{if } h \geqslant 0 \\ &= h\dot{X}(\theta-) \quad \text{if } h \leqslant 0. \end{aligned}$$

It is easy to see that for a fixed $j > 0$

$$\sup_{|h| \leqslant j} |X_{1n}(h)-X_1(h)| \to 0 \text{ as } n \to \infty$$

and weak convergence of $W_n(.)$ as claimed in the theorem follows.

Corollary 3.1. *Let the hypothesis of Theorem* 3.2 *hold and further let* $G(\theta)$ *be a normal r.v. with mean zero and variance* $\sigma^2(\theta) > 0$ *and let* $\dot{X}(\theta+)$ *and* $\dot{X}(\theta-)$ *be both nonzero and of the same sign. Then the hypothesis of Theorem* 3.1 *hold and hence its conclusion with*

$$\begin{aligned} P(Y \leqslant t) &= \Phi(t|\dot{X}(\theta-)|/\sigma(\theta)) \text{ for } t \leqslant 0 \\ &= \Phi(t|\dot{X}(\theta+)|/\sigma(\theta)) \text{ for } t \geqslant 0 \end{aligned}$$

where Φ *is standard normal d f.*

Proof. First we show that (ii) of Theorem 3.1 holds.

Let $\dot{X}(\theta+) > 0$ (and hence $\dot{X}(\theta-) > 0$) If (at a ω) $G(\theta) \geqslant 0$ then $W_n(h) = 0$ has a unique solution namely $h = \frac{G(\theta)}{\dot{X}(\theta-)}$ hence Y (at such a ω) is equal to $-\frac{G(\theta)}{\dot{X}(\theta-)}$. Using similar arguments, we have

$$Y = -\frac{G(\theta)}{\dot{X}(\theta-)} \qquad \text{if} \quad G(\theta)\,\dot{X}(\theta-) \geqslant 0$$

$$= -\frac{G(\theta)}{\dot{X}(\theta+)} \qquad \text{if} \quad G(\theta)\,\dot{X}(\theta+) \leqslant 0$$

Thus (ii) of Theorem 3.1 holds and also $P(Y \leqslant t)$ is as claimed.
Now note that a sample path (at a ω) of $W(h)$ is a line cutting X—axis at $Y(\omega)$. Let $f \in C(-\infty, \infty)$ be such a line and let $\{f_n\}$ be a sequence in $C(-\infty, \infty)$ converging to f (in the topology of u.c.c.) ; it is easy to see that

$$T_i(f_n) \rightarrow T_i(f) \text{ as } n \rightarrow \infty \text{ for } i = 1,2$$

Thus (iii) of Theorem 3.1 holds.

Given a consistent estimate $\hat{\theta}_1$ of θ and set up of Theorem 3.2, the following lemma shows how to get $\hat{\theta}_c$ satisfying (2.2).

Lemma 3.1. *Let the hypothesis of Theorem* 3.2 *hold and further let* $X(t)$ *be monotone in a neighbourhood of* θ *and* $\dot{X}(\theta+)$ *and* $\dot{X}(\theta-)$ *be both nonzero. Then, given a consistent estimate* $\hat{\theta}_1$ *of* θ *we can get a estimate* $\hat{\theta}_c$ *satisfying* (2.2).

Proof. Let $\hat{\theta}_c$ be defined as below

$\hat{\theta}_c$ = solution of $X_n(t) = 0$ nearest to $\hat{\theta}_1$ if $X_n(t) = 0$ for at least one t
$= \hat{\theta}_1$ otherwise.

Fix a $\epsilon > 0$ such that (3.1) holds and assume that $X(t)$ is non-increasing in $[\theta-3\epsilon, \theta+3\,\epsilon]$ so that $\dot{X}(\theta+)$ and $\dot{X}(\theta-)$ are both negative.

Using $W_n(.) \xrightarrow{w} W(.)$ it is easy to see that for given $\eta_1 > 0$ and $\eta_2 > 0$ $\exists\, K > 0$ and n_0 such that for all $n \geqslant n_0$

$$P\left\{\inf_{[-K,K]} W_n(h) < -\eta_1, \sup_{[-K,\,K]} W_n(h) > \eta_1\right\} \leqslant 1 - \eta_2 \qquad \ldots \quad (3.8)$$

Define B_{in} for $i = 1, 2, 3$ and 4 as below

$$B_{1n} = \{X_n(t) \text{ changes sign at least once in } [\theta - n^{-1/2} \log n, \theta + n^{-1/2} \log n]\},$$

$$B_{2n} = \{\theta - \epsilon < \hat{\theta}_1 < \theta + \epsilon\},$$

$$B_{3n} = \left\{ \inf_{[\theta-3\epsilon,\ \theta-n^{-1/2}(\log)^{1/2}]} X_n(t) > 0 \right\} \text{ and}$$

$$B_{4n} = \left\{ \sup_{[\theta+n^{-1/2}(\log n)^{1/2},\ \theta+3\epsilon\}} X_n(t) < 0 \right\}$$

Note that on $\bigcap_{i=1}^{4} B_{in}$ we have

$$|\hat{\theta}_c - \theta| < n^{-1/2} (\log n)^{1/2}.$$

Hence it is enough to show that for $i = 1, 2, 3$ and 4

$$P(B_{in}) \to 1. \qquad \dots \quad (3.9)$$

Note that (3.9) can be proved for $i = 1$ by using (3.8) and also for $i = 2$ by using the consistency of $\hat{\theta}_1$.

Now note that

$$\inf_{[\theta-3\epsilon,\ \theta]} \sqrt{n}(X_n(.) - X(.)) \xrightarrow{w} \inf_{[\theta-3\epsilon,\ \theta]} G(.)$$

hence

$$P\left\{ \inf_{[\theta-3\epsilon,\ \theta-n^{-1/2}(\log n)^{1/2}]} \sqrt{n}(X_n(.) - X(.)) > -(\log\ n)^{1/3} \right\} \to 1$$

and using the decreasing nature of $X(.)$, we have

$$P\left\{ \operatorname{Inf}_{[\theta-3\epsilon,\ \theta-n^{-1/2}(\log n)^{1/2}]} \sqrt{n}\, X_n(h) > \sqrt{n}\, X\, (\theta - n^{-1/2} (\log n)^{1/2}) - (\log n)^{1/3} \right\} \to 1$$

Now (3.9) follows easily for $i = 3$ and using similar arguments also for $i = 4$. Similar arguments applied to $\sqrt{n}(X(.) - X_n(.))$ give us (3.9) in the case when $X(.)$ is non-decreasing there by completing the proof of the lemma.

4. Application to M-estimates

One can apply the results of Section 3 to obtain the asymptotic distribution of M-estimates. In this regard refer to Boos and Serfling (1980) and notations used there. The following theorem is in the spirit of their Theorem 3.2 and can be proved easily using the results of Section 3.

Theorem 4.1. *Let (i)—(v) below hold.*

(i) *$\psi(x, t)$ and $\lambda_F(t)$ are continuous functions of t for t in a neighbourhood of t_0 (here $\lambda_F(t) = \int \psi(x, t)\, dF(x)$ and $\lambda_F(t_0) = 0$)*

(ii) $\int \psi^2(x, t)\, dF(x) < \infty$ *for all* t *in a neighbourhood of* t_0.

(iii) $\lambda_F(t)$, *as a function of* t, *is monotone in a neighbourhood of* t_0 *and* $\lambda'_F(t_0+)$ *and* $\lambda'_F(t_0-)$ *are nonzero.*

(iv) For same $\epsilon > 0$, *as a stochastic process on* $C[t_0-\epsilon, t_0+\epsilon]$, $Y_n(t) \overset{w}{\rightarrow} G(t)$ *where*

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\psi(X_i, t) - \lambda_F(t)),$$

(v) $\lambda_{F_n}(T_n) = 0$ *and* $T_n \overset{p}{\rightarrow} t_0$.

Then

$$\sqrt{n}(T_n - t_0) \overset{w}{\rightarrow} Y \tag{4.1}$$

where

$$\begin{aligned} P(Y \leqslant y) &= \Phi(y\,|\,\lambda'_F(t_0-)\,|\,/\sigma(t_0, F)) \quad \text{for } y \leqslant 0 \\ &= \Phi(y\,|\,\lambda'_F(t_0+)\,|\,/\sigma(t_0, F)) \quad \text{for } y \geqslant 0 \end{aligned}$$

and

$$\sigma^2(t_0, F) = \int \psi^2(x, t_0) dF(x) - \lambda^2_F(t_0).$$

Remark 2. Conditions needed for Theorem 3.2 of Boos and Serfling (1980) and those needed for Theorem 4.1 above are not comparable.

Remark 3. If, instead of (v) of Theorem 4.1, we are given a consistent estimate of t_0 then using Lemma 3.1 one can get a T_n such that $P(\lambda_{F_n}(T_n) = 0) \rightarrow 1$ and (4.1) holds for T_n.

Remark 4. Let F_n denote the continuous version of empirical d.f. and let for $0 < p < 1$, ξ_p and $\hat{\xi}_p$ denote, respectively solutions of $F(x) = p$ and $F_n(x) = p$. If F is continuous in a neighbourhood of ξ_p and $F'(\xi_p+)$ and $F'(\xi_p-)$ are nonzero then using Theorem 4.1 we get asymptotic distribution of $\hat{\xi}_p$.

5. One-step M-estimates

In the set up of M-estimation, given a $\sqrt{n}$- consistent estimate of t_0 one can define a one-step M-estimate of t_0 and obtain its asymptotic distribution (see Serfling, 1980, Chapter 7), Bickel (1975) and Welsh (1988)). One-step procedure can also be developed for the equation (1.2).

Let the conditions of Theorem 3.2 hold and further let $\dot{X}(t)$ be continuously differentiable at θ with $\dot{X}(\theta) \neq 0$.

Let $\hat{\theta}_1$ be such that

$$\sqrt{n}\,(\hat{\theta}_1 - \theta) = O_p(1). \qquad \ldots \quad (5.2)$$

Define a one-step estimate $\hat{\theta}_2$ of θ by

$$0 = X_n(\hat{\theta}_1)+(\hat{\theta}_2-\hat{\theta}_1)\,\dot{X}(\hat{\theta}_1).$$

Note that

$$\sqrt{n}(X_n(\hat{\theta}_1)-X(\hat{\theta}_1))+\sqrt{n}(\hat{\theta}_2-\theta)\,\dot{X}(\theta)\xrightarrow{p} 0. \quad \dots \quad (5.3)$$

Using (3.7) we have for every $K > 0$,

$$\sup_{[-K,K]} W_{1n}(h) \xrightarrow{w} \sup_{[-K,K]} G_1(h) = G(\theta)$$

and

$$\inf_{[-K,K]} W_{1n}(h) \xrightarrow{w} \inf_{[-K,K]} G_1(.) = G(\theta).$$

Hence using (5.2) we have

$$\sqrt{n}(X_n(\hat{\theta}_1)-X(\hat{\theta}_1)) \xrightarrow{w} G(\theta).$$

Now using (5.3) we get

$$\sqrt{n}(\hat{\theta}_2-\theta) \xrightarrow{w} G(\theta)/(-\dot{X}(\theta)).$$

References

Basu, A. P., Ghosh, J. K. and Joshi, S. N. (1988). On estimating change point in a failure rate in *Statistical Decision Theory and Related Topics IV Volume* 2. Ed S. S. Gupta and J. O. Berger, Springer-Verlag, *New York*, 239-252.

→ Bickel, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70**, No. 350, 428-434.

→ Boos, D. D. and Serfling, R. J. (1980). A note on differentials and the CLT and LIL for statistical functions, with application to *M*-estimates. *Ann. Statist*, **8**, No. 3, 618-624.

→ Nguyen, H. T., Rogers, G. S. and Walker, E. A. (1984). Estimation in change-point hazard rate models, *Biometrika*, **71**, 299-405.

Prakasa Rao, B. L. S. (1968). Estimation of the location of the cusp of a continuous density, *Ann. Math. Statist*, **39**, 76-87.

——— (1986). Weak convergence of least squares process in the smooth case, *Statistics*, **17**, 505-516.

Sen, P. K. (1981). *Sequential Nonparametrics*, *Wiley*, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, *Wiley*, *New York*.

→ Welsh, A. H. (1989). On *M*-processes and *M*-estimation, *Ann. Statist*, **17**, No. 1, 337-361.

Statistics and Mathematics Division
Indian Statistical Institute
203 B. T. Road
Calcutta 700 035
India.

Statistics and Mathematics Unit
Indian Statistical Institute
8th Mile, Mysore Road
Bangalore 560 059
India.