Indian J. pure appl. Math., **25**(10): 1017-1029, October 1994

ON THE SPEED OF CONVERGENCE OF LOG-LIKELIHOOD RATIO PROCESSES TO STANDARD MIXTURE OF NORMAL DISTRIBUTIONS

A. K. BASU[1] AND DEBASIS BHATTACHARYA[2]

[1]*Department of Statistics, University College of Science, New Science Building (5th Floor),* 35, *Ballygunge Circular Road, Calcutta* 700 019

[2]*Institute of Agriculture, Visva-Bharathi University, Shantiniketan* 731 235, *West Bengal*

(*Received* 4 *August* 1992; *after revision* 1 *March* 1994; *accepted* 26 *April* 1994)

Under a very general dependence set-up, we obtain results on rate of convergence of log-likelihood ratio processes to standard mixture of normal distributions. Martingale techniques and the principle of contiguity have been exploited throughout. Examples of an explosive autoregressive process of first order and a super-critical Galton-Watson branching process have also been discussed in this context.

INTRODUCTION

It is well known that the statistical information contained in the sample regarding the parameter of interest is described in terms of the likelihood ratios (henceforth LRS) of the sample. A very close relationship between problems of asymptotic likelihood-based inference and the behaviour of log-likelihood ratios, as the number of observations increases, was revealed in LeCam's[12] fundamental work. The statement concerning maximum likelihood estimators can be considered as a statement about the local behaviour of the log-likelihood ratios when the number of observations is large. In LeCam's work and in numerous subsequent works of other authors, it was established that many important properties of statistical estimators and testing of hypotheses follow from the asymptotic normality of log-likelihood ratio for contiguous alternatives. Standard test statistics are usually simple monotonic function of an LR test statistic. The distribution of a log-likelihood ratio statistic provides the best possible power to which the powers of other tests are compared in asymptotic efficiency studies (see Pfanzagl[13]). So it is of prime interest to study the asymptotic behaviour of a LRS and its related statistics (that is, statistics which are functionals of LR statistics) from the view point of asymptotic theory of inference. Various studies by various authors, regarding asymptotic distribution of log-likelihood ratio statistics, have been made. Asymptotic normality of the log-likelihood function for a general stochastic process was established by Roussas[14]. Jeganathan[10] shows that for dependent observations the sequence of log-LRS is locally asymptotically mixed normal (LAMN).

In many situations, the full likelihood may not be available, or, if available, is very complicated in nature. There, standard statistical practice is to use a good approximation to the likelihood which is, generally, expressed in terms of the limiting likelihood. Such practice will not be very worthwhile, in general (for practical purposes), unless the rate of convergence of the limiting distribution to an exact distribution is very high. In a recent paper of Basu and Bhattacharya[4], the rate of convergence of a normalized log-likelihood ratio statistic to a standard Normal variable, for a LAN family of distributions, has been explored. In this paper, we try to extend our effort in finding the rate of convergence of a log-LR statistic to a standard mixture of Normal distributions, for a LAMN family of distributions. Martingale techniques and the principle of contiguity play an important role and have been exploited in this paper, as and when required.

In section 1, all necessary notation and assumptions used throughout the paper are introduced. A series of lemmas required to prove the main theorem are stated in section 2. Section 3 is fully devoted to the main theorem and its proof. Finally, section 4 contains two pertinent examples, and section 5 contains remarks and discussion regarding the rates of convergence.

1. Notation and Assumptions

Let $X_1, X_2, ..., X_n$ be the first n random variables from a certain stochastic process. Let these r.v.'s be defined on the probability space $(x, \mathcal{F}, P_\theta)$ and take values in the space (S, ζ), where S is a Borel subset of a Euclidean space and ζ is the σ-field of Borel subsets of S. It is assumed that the joint probability law of any finite set of such r.v.'s has some known functional form except for the unknown parameter θ involved in the distribution, where θ lies in an open subset Θ of R^k, $k \geq 1$.

Let $\mathcal{F}_n = \sigma(X_1, X_2, ..., X_n)$ be the σ-field induced by the r.v.'s $(X_1, X_2, ..., X_n)$ and assume that for $j \geq 2$, a regular conditional probability measure of X_j, given $(X_1 = x_1, X_2 = x_2, ..., X_{j-1} = x_{j-1})$ is absolutely continuous with respect to a σ-finite measure μ_j with corresponding density $f_j(x_j \mid x_1, ..., x_{j-1}; \theta)$, and the probability measure of X_1 is absolutely continuous with respect to a σ-finite measure μ_1 with corresponding density $f_1(x_1 \mid \theta)$. For the sake of notational simplicity, we write $f_j(x_j \mid x_1, ..., x_{j-1}; \theta) = f_j(\theta)$, for $j \geq 2$, and $f_1(x_1 \mid \theta) = f_1(\theta)$. Suppose θ_0 is the true value of the parameter θ and let $\theta_n^* = \theta^* = \theta_0 + \partial_n h$ be other values of $\theta; \theta_0$, $\theta_n^* \in \Theta$, where $h \in R^k$ and $\partial_n^{k \times k}$ is a positive definite matrix, which may depend on θ_0 but is independent of the observations and such that the matrix norm $\| \partial_n \|$, defined by the square root of the sum of squares of all elements, tends to zero as $n \to \infty$. One way of selecting ∂_n is to set

$$\partial_n' \, \partial_n = \left[\sum_{j=1}^{n} E\left[\eta_j(\theta_0)\, \eta_j'(\theta_0)\right] \right]^{-1},$$

where $$\eta_j^{k \times 1}(\theta_0) = \left[\frac{\partial}{\partial \theta} \log f_j(\theta) \right]_{\theta = \theta_0}.$$

In the i.i.d. case, one may set $\partial_n' \, \partial_n = n^{-1} I_k$, where I_k is a unit matrix of order k (c.f. Jeganathan[10]). Let $P_{n\theta}$ be the restriction of P_θ to $\mathcal{F}_n$. We can write the likelihood of $P_{n\theta_n^*}$ with respect to $P_{n\theta_0}$ as

$$L_n(\theta_n^*, \theta_0) = \frac{dP_{n\theta_n^*}}{dP_{n\theta_0}}$$

$$= \frac{\prod_{j=1}^{n} f_j(\theta_n^*)}{\prod_{j=1}^{n} f_j(\theta_0)}. \qquad \text{... (1.1)}$$

So the log-likelihood $\Delta_n(\theta_n^*, \theta_0)$ can be written as

$$\Delta_n(\theta_n^*, \theta_0) = \log L_n(\theta_n^*, \theta_0)$$

$$= \sum_{j=1}^{n} \log \frac{f_j(\theta_n^*)}{f_j(\theta_0)}. \qquad \text{... (1.2)}$$

In the sequel, for any vector $y \in R^k$, the norm of y is $|y| = \left(\sum_{i=1}^{k} y_i^2 \right)^{1/2}$, equality and inequality between vectors are to be understood elementwise and those between r.v.'s are to be understood in almost sure sense. Unless otherwise stated, expectation of a random variable is to be understood under θ_0.

Assumptions — Here we make the following set of assumptions :

(A1) For almost all $(x_1, x_2, ..., x_j)$ and for every $j \geq 1$, the function $\theta \to f_j(\theta)$ is absolutely continuous in θ.

(A2) For $\prod_{i=1}^{j} \mu_i$ almost all $(x_1, x_2, ..., x_j)$, $j \geq 1$ and for every $\theta \in \Theta$, the quantity

$$\dot{f}_j(\theta) = \frac{\partial}{\partial \theta} f_j(\theta) = \left(\frac{\partial}{\partial \theta_1}, ..., \frac{\partial}{\partial \theta_k} \right)' f_j(\theta) \text{ exists.}$$

Define

$$\left.\begin{aligned} \dot{\xi}_j(\theta) &= \frac{\dot{f}_j(\theta)}{f_j^{1/2}(\theta)}, \text{ if } \dot{f}_j(\theta) \text{ exists and } f_j(\theta) > 0 \\ &= 0, \quad \text{otherwise,} \end{aligned}\right] \qquad \text{... (1.3)}$$

and $\eta_j(\theta) = \dot{\xi}_j(\theta)/f_j^{1/2}(\theta)$.

(A3) For every $h \in R^k$ and $\theta \in \Theta$,

$$E_{\theta_0}\left[\int |h' \partial_n \dot{\xi}_j(\theta)|^2 d\mu_j\right] < \infty, \ 1 \le j \le n < \infty.$$

(A4) For every $h \in R^k$ and for every $\theta \in \Theta$,

$$\sup_{a \le h \le b} \sum_{j=1}^{n} E_{\theta_0}\left\{\int |h' \partial_n [\dot{\xi}_j(\theta_n^*) - \dot{\xi}_j(\theta_0)]|^2 d\mu_j\right\} \to 0$$

as $n \to \infty$ for some $a < 0$ and $b > 1$.

(A5) $E[\eta_j(\theta_0) \mid \mathcal{F}_{j-1}] = 0$ for every $j \ge 1$.

(A6) There exists an a.s. positive definite random matrix $T(\theta_0)$ such that

$$\partial_n \sum_{j=1}^{n} E[\eta_j(\theta_0)\eta_j'(\theta_0) \mid \mathcal{F}_{j-1}]\, \partial_n - T(\theta_0)$$

converges to zero in probability.

(A7) For every $\varepsilon > 0$ and $h \in R^k$,

$$\sum_{j=1}^{n} E[|h' \partial_n \eta_j(\theta_0)|^2 I(|h' \partial_n \eta_j(\theta_0)| > \varepsilon)] \to 0,$$

where $I(C)$ stands for the indicator of the set C.

(A8) For every $h \in R^k$, there exists a constant $M > 0$ such that

$$\sup_{n \ge 1} \sum_{j=1}^{n} E[|h' \partial_n \eta_j(\theta_0)|^2] \le M .$$

(A9) For some constants ε_n (which $\downarrow 0$ as $n \to 0$) and for sufficiently large n,

$$E_{\theta_0}\left\{\sup_{|h' \partial_n| \le \varepsilon_n} \left\| \partial_n \sum_{j=1}^{n} [\eta_j(\theta_n^{**}) - \eta_j(\theta_0)] \right\| \right\} \le C\, \varepsilon_n^{1/2},$$

where $|\theta_n^{**} - \theta_0| \le |\partial_n h|$ (elementwise).

(A10) All 3rd order moments of the random matrix $T(\theta_0)$ appearing in (A6) are finite.

2. Some Technical Lemmas

Lemma 2.1 (Hall and Heyde[8], p. 396) — Let $\left\{S_{ni} = \sum_{j=1}^{i} X_{nj},\ \mathcal{F}_{ni},\ 1 \le i \le k_n\right\}$ be a zero mean martingale for each $n \ge 1$. Suppose that martingale differences satisfy the following conditions :

(i) for all $\varepsilon > 0$, $\sum_{i=1}^{k_n} E[X_{ni}^2 I(|X_{ni}| > \varepsilon)] \to 0$.

(ii) $\sum_{i=1}^{k_n} X_{ni}^2 \xrightarrow{P} \eta^2$,

where the random variable $\eta > 0$ a.s. .

(iii) $\mathcal{F}_{ni} \subseteq \mathcal{F}_{n+1,\,i}$ for all $i \le k_n$.

Then $P(S_{nk_n} \le x) \to E[\Phi(\eta^{-1} x)]$ for all real x, where Φ is the standard normal distribution function.

The above result also holds if condition (ii) is replaced by

(ii)′ $\sum_{i=1}^{k_n} E(X_{ni}^2 | \mathcal{F}_{n,\,i-1}) \xrightarrow{P} \eta^2$.

Lemma 2.2 — Let $\left\{ S_{ni} = \sum_{j=1}^{i} X_{nj},\ \mathcal{F}_{ni},\ 1 \le i \le n \right\}$ be a zero mean square integrable martingale for each $n \ge 1$. Set $S_n = S_{nn}$, $S_0 = 0$,

$$U_i^2 = \sum_{j=1}^{i} X_{nj}^2 \text{ and } V_i^2 = \sum_{j=1}^{i} E(X_{nj}^2 | \mathcal{F}_{n,j-1}),\ 1 \le i \le n,$$

with $U_0 = 0$. Suppose $0 \le m \le n$ and define

$$L_{nm} = E(U_m^{3/2}) + \sum_{j=m+1}^{n} E|X_{nj}|^3 + E|U_n^2 - \eta^2|^{3/2} + E|\eta^2 - E(\eta^2 | \mathcal{F}_{nm})|^{3/2},$$

where η is a random variable as defined in Lemma 2.1, $\mathcal{F}_{nm} = \sigma(X_{n1}, X_{n2}, \ldots, X_{nm})$, and $\mathcal{F}_{n0}$ is the trivial σ-field. Assume that $E[\eta^3] < C$, where C is a positive constant. Then there exists a constant A such that, for all x and m and whenever $L_{nm} \le 1$,

$$|P(S_n \le x) - E[\Phi(\eta^{-1} x)]| \le A L_{nm}^{1/4} (1 + |x|^{9/4})^{-1}.$$

PROOF : Follows directly from Theorem 1 of Hall and Heyde[8], with $\partial = 1/2$.

Lemma 2.3 — Under the set of notation and assumptions of Lemma 2.2, there exists a constant A such that, for all x and m,

$$\sup_x |P(S_n \le x) - E[\Phi(\eta^{-1} x)]| \le A L_{nm}^{1/4}.$$

PROOF : Follows from the Corollary to Theorem 1 of Hall and Heyde[8], with $\partial = 1/2$.

Remarks : (i) The result of Lemmas 2.2 and 2.3 also holds when $E\,|\,U_n^2-\eta^2\,|^{3/2}$ in the definition of L_{nm} is replaced by $E\,|\,V_n^2-\eta^2\,|^{3/2}$.

(ii) Both lemmas will hold when $m = 0$ and for general ∂, $0 \le \partial \le 1$ as in Hall and Heyde[8] with appropriate modification in definitions and moment conditions.

Lemma 2.4 — For $m = 0, 1, 2, ..., n$, set,

$$Q_{n1}(m) = E\left(\left|\sum_{j=1}^{m} X_{nj}^2\right|^{3/2}\right) + \sum_{j=m+1}^{n} E\,|\,X_{nj}|^3 + E\left(\left|\sum_{j=m+1}^{n} E(X_{nj}^2\,|\,\mathcal{F}_{n,j-1}) - \eta_{nm}^2\right|^{3/2}\right) + E\,(|\,\eta^2-\eta_{nm}^2\,|^2),$$

where $\mathcal{F}_{ni} = \sigma(X_{n1}, X_{n2}, ..., X_{ni})$, with $\mathcal{F}_{n0}$ being the trivial σ-field, and $\eta_{nm}^2 = E(\eta^2\,|\,\mathcal{F}_{nm})$. Assume that $E[\eta^4] < \infty$, and $E[\eta^{-4}] < \infty$. Then there exists a constant C such that, for all $x \in R$ and $m = 0, 1, 2, .., n-1$ whenever $Q_{n1}(m) \le 1$,

$$|\,P(S_n \le x) - E[\Phi(\eta^{-1}x)]\,| \le C\,Q_{n1}^{1/4}(m)\,(1 + |\,x\,|^3)^{-1}.$$

PROOF : Follows from Haeusler and Joos[6] (p. 1701), with the choice of $\partial = 1/2$.

3. MAIN THEOREM

From (1.2), we find that

$$\Delta_n(\theta_n^*, \theta_0) = \sum_{j=1}^{n} \log \frac{f_j(\theta_n^*)}{f_j(\theta_0)}\,.$$

It is well known that the sequence of likelihood ratios forms a martingale [the correct hypothesis being represented in the denominator and the distributions being mutually absolutely continuous; e.g. Basawa and Prakasa Rao[1] (pp. 363 and 392)] and hence the sequence of log-likelihood ratios is a super-martingale. To handle $\Delta_n(\theta_n^*, \theta_0)$, we require a suitable breaking of it in terms of a martingale which we can achieve nicely through the application of mean value theorem in the following way :

$$\Delta_n(\theta_n^*, \theta_0) = \sum_{j=1}^{n} [\log f_j(\theta_n^*) - \log f_j(\theta_0)]$$

$$= \sum_{j=1}^{n} h'\,\partial_n\,\eta_j(\theta_n^{**}), \text{ where } |\,\theta_n^{**} - \theta_0\,| \le |\,\partial_n h\,|$$

(elementwise)

and $\eta_j(\theta_0)$ is satisfied in section 1, and this is equal to :

$$= \sum_{j=1}^{n} h' \,\partial_n \,\eta_j(\theta_0) + \sum_{j=1}^{n} h' \,\partial_n \,[\eta_j(\theta_n^{**}) - \eta_j(\theta_0)]$$

$$= \sum_{j=1}^{n} Y_{nj} + \sum_{j=1}^{n} Z_{nj}, \qquad \text{... (3.1)}$$

where $Y_{nj} = h' \,\partial_n \,\eta_j(\theta_0)$, $\{Y_{nj}\}$ is a martingale difference sequence and $Z_{nj} = h' \,\partial_n \,[\eta_j(\theta_n^{**}) - \eta_j(\theta_0)]$ for each $n \geq 1$.

We note that as $n \to \infty$, $\|\partial_n\| \to 0$, so that $\theta_n^{**} \to \theta_0$. We see that (3.1) satisfies Doob's decomposition of a super-martingale (Hall and Heyde[7], p. 51).

Theorem 3.1 — Under contiguous alternatives and the set of assumptions and notation stated in sections 1 and 2, we have the following non-uniform error bound

$$\left| P_{\theta_0}(\Delta_n(\theta_n^*, \theta_0) \leq x) - E_{\theta_0}\left[\Phi\left(\frac{x + \frac{1}{2} h' \,T(\theta_0) h}{\sqrt{(h' \,T(\theta_0) h)}} \right) \right] \right|$$

$$\leq O(\varepsilon_n^{1/4}) + A L_{nm}^{1/4} (1 + |x|^{9/4})^{-1},$$

where Φ is the distribution function of the standard normal distribution.

Theorem 3.2 — Under contiguous alternatives and the set of assumptions and notation stated in sections 1 and 2, we have the following uniform error bound

$$\sup_x \left| P_{\theta_0}(\Delta_n(\theta_n^*, \theta_0) \leq x) - E_{\theta_0}\left[\Phi\left(\frac{x + \frac{1}{2} h' \,T(\theta_0) h}{\sqrt{(h' \,T(\theta_0) h)}} \right) \right] \right|$$

$$\leq O(\varepsilon_n^{1/4}) + A L_{nm}^{1/4}.$$

Theorem 3.3 — Under the same set of assumptions and notation stated in sections 1 and 2, except that assumption (A10) is replaced by (A10)' namely : All 4th order moments of $T(\theta_0)$ and $T^{-1}(\theta_0)$ are finite, we have the following non-uniform error bound

$$\left| P_{\theta_0}(\Delta_n(\theta_n^*, \theta_0) \leq x) - E_{\theta_0}\left[\Phi\left(\frac{x + \frac{1}{2} h' \,T(\theta_0) h}{\sqrt{(h' \,T(\theta_0) h)}} \right) \right] \right|$$

$$\leq O(\varepsilon_n^{1/4}) + C\, Q_{n1}^{1/4}(m)\,(1 + |x|^3)^{-1},$$

where $Q_{n1}(m)$ is given in Lemma 2.4.

Proof of Theorem 3.1 : From (3.1), we have

$$\left| P_{\theta_0}(\Delta_n(\theta_n^*, \theta_0) \le x) - E_{\theta_0}\left[\Phi\left(\frac{x + \frac{1}{2} h' T(\theta_0)h}{\sqrt{(h' T(\theta_0)h)}}\right)\right]\right|$$

$$= \left| P_{\theta_0}\left(\sum_{j=1}^{n} Y_{nj} + \sum_{j=1}^{n} Z_{nj} \le x\right) - E_{\theta_0}\left[\Phi\left(\frac{x + \frac{1}{2} h' T(\theta_0)h}{\sqrt{(h' T(\theta_0)h)}}\right)\right]\right|$$

$$\le \left| P_{\theta_0}\left(\sum_{j=1}^{n} Y_{nj} \le x\right) - E_{\theta_0}\left[\Phi\left(\frac{x + \frac{1}{2} h' T(\theta_0)h}{\sqrt{(h' T(\theta_0)h)}}\right)\right]\right|$$

$$+ P_{\theta_0}\left(\left|\sum_{j=1}^{n} Z_{nj}\right| \ge \varepsilon_n\right) + O(\varepsilon_n)$$

$$= I + II + O(\varepsilon_n), \qquad \text{... (3.2)}$$

for any sequence of positive constants $\{\varepsilon_n\}$, $n \ge 1$, tending to zero. Now

$$II = P_{\theta_0}\left[\left|\sum_{j=1}^{n} Z_{nj}\right| \ge \varepsilon_n\right]$$

$$= P_{\theta_0}\left[\left|\sum_{j=1}^{n} h' \partial_n \{\eta_j(\theta_n^{**}) - \eta_j(\theta_0)\}\right| \ge \varepsilon_n\right]$$

$$\le P_{\theta_0}\left[\sup_{|h' \partial_n| \le \varepsilon_n} \left| h' \partial_n \sum_{j=1}^{n} \{\eta_j(\theta_n^{**}) - \eta_j(\theta_0)\}\right| \ge \varepsilon_n\right]$$

$$\le \frac{E_{\theta_0}\left[\sup_{|h' \partial_n| \le \varepsilon_n} \left| h' \partial_n \sum_{j=1}^{n} \{\eta_j(\theta_n^{**}) - \eta_j(\theta_0)\}\right|\right]}{\varepsilon_n}$$

(using Markov inequality)

$$\le \|\partial_n^{-1}\| E_{\theta_0}\left[\sup_{|h' \partial_n| \le \varepsilon_n} \left| \partial_n \sum_{j=1}^{n} \{\eta_j(\theta_n^{**}) - \eta_j(\theta_0)\}\right|\right]$$

$$\le \|\partial_n^{-1}\| C\, \varepsilon_n^{1/2} \qquad \text{(due to (A9))} \qquad \text{... (3.3)}$$

$$\le O(\varepsilon_n^{1/4}) \qquad \text{(selecting } \|\partial_n^{-1}\| = \varepsilon_n^{-1/4}\text{).} \qquad \text{... (3.4)}$$

Also

$$I = \left| P_{\theta_0}\left(\sum_{j=1}^{n} h' \partial_n \eta_j(\theta_0) \le x \right) - E_{\theta_0}\left[\Phi\left(\frac{x + \frac{1}{2} h' T(\theta_0) h}{\sqrt{(h' T(\theta_0) h)}} \right) \right] \right|$$

$$\le \left| P_{\theta_0}\left(\sum_{j=1}^{n} h' \partial_n \eta_j(\theta_0) \le x \right) - E_{\theta_0}\left[\Phi\left(\frac{x}{\sqrt{(h' T(\theta_0) h)}} \right) \right] \right|$$

$$+ E_{\theta_0}\left| \Phi\left(\frac{x}{\sqrt{(h' T(\theta_0) h)}} \right) - \Phi\left(\frac{x}{\sqrt{(h' T(\theta_0) h)}} + \frac{1}{2}\sqrt{(h' T(\theta_0) h)} \right) \right|$$

$$\le A L_{nm}^{1/4} (1 + |x|^{9/4})^{-1} + E_{\theta_0}\left[\frac{1}{2}\sqrt{(h' T(\theta_0) h)} \right]$$

(using Lemmas 2.1 and 2.2)

$$\le A L_{nm}^{1/4} (1 + |x|^{9/4})^{-1} + \| \partial_n^{-1} \| \varepsilon_n$$

$$\le A L_{nm}^{1/4} (1 + |x|^{9/4})^{-1} + O(\varepsilon_n^{3/4}) \text{ (selecting } \| \partial_n^{-1} \| = \varepsilon_n^{-1/4} \text{ and due to assumption (A10)).} \quad \text{... (3.5)}$$

So we have that the left-hand side of (3.2) is

$$\le O(\varepsilon_n^{1/4}) + A L_{nm}^{1/4} (1 + |x|^{9/4})^{-1}. \quad \text{... (3.6)}$$

Proof of Theorem 3.2 : In the 1st part of (3.5), we apply Lemmas 2.1 and 2.3 and obtain the result of Theorem 3.2.

Proof of Theorem 3.3 : In the 1st part of (3.5), we apply Lemmas 2.1 and 2.4 and obtain the desired result.

4. Examples

Example 4.1 — An explosive auto-regressive process of first order.

We consider a first order auto-regressive process $\{X_j,\ j = 1, 2, ...\}$ defined by

$$X_j = \theta X_{j-1} + \varepsilon_j,\ X_0 = 0,$$

where ε_j's are i.i.d. $N(0, 1)$ random variables. We consider $|\theta| > 1$ which refers to regular non-ergodic (explosive) case. Two of our previous papers[3, 5] include discussion on this model.

For this model, we notice that,

$$T_n(\theta_0) = \frac{(\theta_0^2 - 1)^2}{\theta_0^{2n}} \sum_{j=1}^{n} X_{j-1}^2 \to \chi_1^2 = T(\theta_0) \text{ a.s. .}$$

So, clearly, assumption (A10) is satisfied in this case. Now we check that for this model (A9) also holds. Here

$$\left| \partial_n \sum_{j=1}^{n} [\eta_j(\theta_n^{**}) - \eta_j(\theta_0)] \right|$$

$$= \left| \partial_n \sum_{j=1}^{n} [X_{j-1}(X_j - \theta_n^{**} X_{j-1}) - X_{j-1}(X_j - \theta_0 X_{j-1})] \right|$$

(where θ_n^{**} is such that $|\theta_n^{**} - \theta_0| \le |\partial_n h|$ and

$\eta_j(\theta_0) = X_{j-1}(X_j - \theta_0 X_{j-1}) = X_{j-1}\varepsilon_j$)

$$= \left| \partial_n \sum_{j=1}^{n} (\theta_0 - \theta_n^{**}) X_{j-1}^2 \right|$$

$$\le \sum_{j=1}^{n} \partial_n |\theta_0 - \theta_n^{**}| X_{j-1}^2 \le \sum_{j=1}^{n} (\partial_n)^2 |h| X_{j-1}^2.$$

So

$$E_{\theta_0}\left\{ \sup_{|h'\partial_n| \le \varepsilon_n} \left| \partial_n \sum_{j=1}^{n} [\eta_j(\theta_n^{**}) - \eta_j(\theta_0)] \right| \right\}$$

$$\le E_{\theta_0}\left\{ \sup_{|h'\partial_n| \le \varepsilon_n} (\partial_n)^2 |h| \sum_{j=1}^{n} X_{j-1}^2 \right\}$$

$$\le \varepsilon_n \partial_n^{-1} E_{\theta_0}\left[\frac{(\theta_0^2 - 1)^2}{\theta_0^{2n}} \sum_{j=1}^{n} X_{j-1}^2 \right]$$

$$\le C \varepsilon_n^{3/4} \le C \varepsilon_n^{1/2}.$$

[For this model, $\partial_n = \dfrac{(\theta_0^2 - 1)}{\theta_0^n}$, $E_{\theta_0}\left[\sum_{j=1}^{n} X_{j-1}^2 \right] = \dfrac{\theta_0^{2n}}{(\theta_0^2 - 1)^2}$ (c.f. Basawa and Scott[2]) and ε_n may be selected to be $\varepsilon_n = \dfrac{(\theta_0^2 - 1)^4}{\theta_0^{4n}}$].

Hence all the results of Theorems 3.1 and 3.2 on the rate of convergence discussed in section 2 can be applied to this model.

Example 4.2 — A super-critical Galton-Watson branching process.

Let $X_0 = 1$, and $X_1, X_2, \ldots, X_n$ be the successive generation sizes of a Galton-Watson branching process with geometric offspring distribution

$$P(X_1 = j) = \theta^{-1}(1 - \theta^{-1})^{j-1}, \; j = 1, 2, \ldots .$$

We assume that the process is super-critical, i.e. $\theta > 1$. Here $E(X_1) = \theta$ and $V(X_1) = \sigma^2(\theta) = \theta(\theta - 1)$. We also assume $\sigma^2(\theta) < \infty$. One of our previous papers (Basu and Bhattacharya[3], pp. 157-158) includes detail discussion on this process.

Let $E^c = \{w \mid X_n(w) > 0, \text{ for all } n\}$ be the set of non-extinction. Conditional on the set of non-extinction, $X_n \to \infty$ a.s. as $n \to \infty$, and in that sense the process is explosive. For simplicity, we suppose $p(0, \theta) = 0$, where $P(x, \theta) = P(X_1 = x; \theta)$ so that all the limit results which are conditional on E^c hold.

For this model, we have the following observations :

(O1) $f_j(\theta) \propto \left(1 - \frac{1}{\theta}\right)^{x_j - x_{j-1}} \left(\frac{1}{\theta}\right)^{x_{j-1}}$.

(O2) $\eta_j(\theta) = \frac{1}{\theta(\theta-1)}(X_j - \theta X_{j-1})$, $\xi_j(\theta) = \eta_j(\theta) f_j^{1/2}(\theta)$.

(O3) $E_\theta\left(\sum_{j=1}^{n} X_{j-1}\right) = \frac{\theta^n - 1}{(\theta - 1)}$ and $E_\theta\left(\sum_{j=1}^{n} X_j\right) = \frac{\theta(\theta^n - 1)}{(\theta - 1)}$.

(O4a) $\frac{(\theta-1)}{\theta^{n+1}} \sum_{j=1}^{n} X_j \to T(\theta)$ a.s. as $n \to \infty$,

(O4b) $T_n(\theta) = \frac{(\theta-1)}{\theta^n} \sum_{j=1}^{n} X_{j-1} \to T(\theta)$ a.s. as $n \to \infty$,

where $T(\theta)$ is a non-negative random variable. Particularly, $T(\theta)$ is a negative exponential random variable with mean unity.

For this process, all the assumptions stated in section 2 have been verified. Assumption (A10), clearly, holds for this process due to (O4b). Here we will only verify that assumption (A9) holds for this process and prefer to omit details of calculations.

Here $\left|\partial_n \sum_{j=1}^{n} [\eta_j(\theta_n^{**}) - \eta_j(\theta_0)]\right|$ (where $|\theta_n^{**} - \theta_0| \le |\partial_n h|$ and $\partial_n = \frac{\theta_0^{1/2}(\theta_0 - 1)}{\theta_0^{n/2}}$ is a scalar quantity)

$$= \left|\partial_n \sum_{j=1}^{n} \left[\frac{X_j - \theta_n^{**} X_{j-1}}{\theta_n^{**}(\theta_n^{**} - 1)} - \frac{X_j - \theta_0 X_{j-1}}{\theta_0(\theta_0 - 1)}\right]\right|$$

$$= \left|\partial_n \sum_{j=1}^{n} \left[\frac{1}{\theta_0(\theta_0 - 1)\theta_n^{**}(\theta_n^{**} - 1)} \{(\theta_0^2 - \theta_n^{**2}) X_j \right.\right.$$

$$\left.\left. + (\theta_n^{**} - \theta_0) X_j + \theta_n^{**} \theta_0 (\theta_n^{**} - \theta_0) X_{j-1}\}\right]\right|$$

$$= \left| \frac{\partial_n}{\theta_0(\theta_0-1)\,\theta_n^{**}(\theta_n^{**}-1)} \sum_{j=1}^{n} \Big[(\theta_0-\theta_n^{**})\,\theta_0 X_j + (\theta_0-\theta_n^{**})(\theta_n^{**}-1) X_j + \theta_n^{**}\,\theta_0(\theta_n^{**}-\theta_0) X_{j-1} \Big] \right|$$

$$= \left| \frac{\partial_n}{\theta_0(\theta_0-1)} \sum_{j=1}^{n} \left[\frac{(\theta_0-\theta_n^{**})\,\theta_0}{\theta_n^{**}(\theta_n^{**}-1)} (X_j - \theta_n^{**} X_{j-1}) + \frac{(\theta_0-\theta_n^{**})}{\theta_n^{**}} X_j \right] \right|,$$

since, using (O4a) and (O4b), we have $\frac{(\theta-1)}{\theta^{n+1}} \sum_{j=1}^{n} (X_j - \theta X_{j-1}) \to 0$ a.s. as $n \to \infty$, the last expression becomes

$$\le C_1 \frac{\partial_n}{\theta_0(\theta_0-1)} \frac{|\theta_0-\theta_n^{**}|}{\theta_n^{**}} \sum_{j=1}^{n} X_j ,$$

where C_1 is a generic constant.

Therefore, using (O3), and the last inequality

$$E_{\theta_0}\left\{ \sup_{|h'\partial_n| \le \varepsilon_n} \left| \partial_n \sum_{j=1}^{n} [\eta_j(\theta_n^{**}) - \eta_j(\theta_0)] \right| \right\}$$

$$\le C_1 \frac{\partial_n\,\varepsilon_n}{(\theta_0-1)\,\theta_n^{**}} \frac{\theta_0^n}{(\theta_0-1)}$$

$$\le C_2\, \partial_n^{-1}\, \varepsilon_n, \text{ since } \partial_n^{-2}(\theta_0) = \frac{\theta_0^n}{\theta_0(\theta_0-1)^2}$$

$$\le C_2\, \varepsilon_n^{1/2}, \text{ choosing } \partial_n = \varepsilon_n^{1/4}.$$

Hence, all the results of Theorems 3.1 and 3.2 on the rate of convergence discussed in section 2 can be applied to this model.

Here, we consider two widely used LAMN models. Regression with integrated processes, Pure-birth processes, and Diffusion processes are some interesting processes with LAMN likelihood ratios where our results can be applied.

5. Remarks and Discussion Regarding Rates of Convergence

So far as the log-likelihood ratio statistic is concerned, this is the only result available regarding the rate of convergence of the log-likelihood ratio statistic to a mixture of normal distributions. Faster rates of convergence may be obtained using C.L.T. for martingale differences under more stringent conditions (viz. Ibragimov[9], Kato[11]). However, in the field of applications, it is often difficult to check even the basic sufficient conditions of those theorems. For this reason, we have used results of Hall and Heyde[8] and Haeusler and Joos[6] in our context.

The rate depends on L_{nm} and $Q_{n1}(m)$, where both of them are functions defined on the process. The rate becomes proper (non-trivial) when both the terms tend to zero as $n \to \infty$. So it may happen that, in practice, a rate better than $\varepsilon_n^{1/4}$ can be achieved. It is worth mentioning that the result of Hall and Heyde[8] is based on the Skorokhod embedding technique which never gives a bound better than $n^{-1/4}$.

Theorem 3.1 and Theorem 3.3 are not directly comparable, since L_{nm}-terms of Theorem 3.1 and $Q_{n1}(m)$ of Theorem 3.3 are different. However, they do provide a proper (non-trivial) rate of convergence, since under the corresponding conditions for weak convergence of martingales to a mixture of normal distributions, $Q_{n1}(m)$ (in Lemma 2.4) goes to zero as $n \to \infty$ for an appropriate choice of $m = m(n)$ (see Haeusler and Joos[6], pp. 1700-1702), and also L_{nm}-terms (in Lemma 2.2) goes to zero as $n \to \infty$ (see Hall and Heyde's[8] discussion concerning L_{nm}-terms, p.396). Then Theorem 3.3 provides a slightly faster non-uniform rate of convergence compared to Theorem 3.1 under slightly higher moment condition on the mixing variables. Further, we note that the bound in Theorem 3.3 contains the same optimal term involving x as Haeusler and Joos (see discussion in Haeusler and Joos[6], pp. 1700-1702).

ACKNOWLEDGEMENT

Thanks to the referees for their suggestions which led to an improvement of the paper.

REFERENCES

1. I. V. Basawa and B. L. S. Prakasa Rao, *Statistical Inference for Stochastic Processes*, Academic Press, 1980.
2. I. V. Basawa and D. J. Scott, *Lecture Notes in Statistics*, Springer-Verlag, New York., 1983.
3. A. K. Basu and D. Bhattacharya, *Calc. Stat. Assoc. Bull.* **37** (1988), 143-59.
4. A. K. Basu and D. Bhattacharya, On the speed of convergence in the Central Limit Theorem of log-likelihood ratio processes. *Technical Report. No.* 3/91, Department of Statistics, Calcutta University. To appear in *J. Theoretical Probab.* **6**(4) (1993).
5. A. K. Basu and D. Bhattacharya, *Cal. Stat. Assoc. Bull.* **39** (1990), 137-49.
6. E. Haeusler and K. Joos, *Ann. Probab.* **61** (1988), 1699-1720.
7. P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*, Academic Press, 1980.
8. P. Hall and C. C. Heyde, *Ann. Probab.* **9** (1981), 395-404.
9. I. A. Ibragimov, *Theory Probab. Appl.* **8** (1963), 83-89.
10. P. Jeganathan, *Sankhyā* **44**, pt.2, Ser-A (1982), 173-212.
11. Y. Kato, *Bull. Math. Statist.* **18** (1979), 1-8.
12. L. LeCam, Locally asymptotically normal families of distributions, *Univ. of California publication in Statistics*, Vol. 3 (1960), pp. 37-98.
13. J. Pfanzagl, Asymptotic expansion in parametric statistical theory. *Developement in Statistics*, Vol.3, Academic Press, New York, 1980, pp. 1-97.
14. G. G. Roussas, *Z. Wahrsch. Verw. Gebiete* (1979), pp. 31-46.