CRRAO Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS)

Research Report

Author (s): **B.L.S. Prakasa Rao**

Title of the Report: **MODERATE DEVIATION PRINCIPLE FOR MAXIMUM LIKELIHOOD ESTIMATOR FOR MARKOV PROCESSES**

Research Report No.: **RR2016-08**

Date: **October 24, 2016**

Prof. C R Rao Road, University of Hyderabad Campus,
Gachibowli, Hyderabad-500046, INDIA.
www.crraoaimscs.org

MODERATE DEVIATION PRINCIPLE FOR MAXIMUM LIKELIHOOD ESTIMATOR FOR MARKOV PROCESSES

B.L.S. PRAKASA RAO
CR Rao Advanced Institute of Mathematics, Statistics and Computer Science, Hyderabad 500046, India

Abstract: After a short review of the properties of the maximum likelihood estimator for discrete time Markov processes, we obtain a moderate deviation result for such an estimator under some regularity conditions using the Gärtner-Ellis theorem for random processes.

Mathematics Subject Classification: 62M05, 62F12, 60F99.

Keywords: Moderate deviation; Maximum likelihood estimation; Markov process.

1 Introduction

Let $\{X_n, n \geq 1\}$ be a stochastic process defined on a probabilty space $(\Omega, \mathcal{B}, P_\theta)$ taking values in a measurable space $(X, \mathcal{F}_X)$. We assume that the parameter $\theta \in \Theta \subset R$ but is unknown. Suppose we observe a sample $(X_1, \ldots, X_n)$ of the process. The problem of estimation of the parameter θ based on the observation $(X_1, \ldots, X_n)$ has been discussed in the literature over the last several years. For instance, see Billingsley (1961) and Prakasa Rao (1972, 1973, 1977) for the case of the discrete time Markov processes and Basawa and Prakasa Rao (1980) and Grenander (1981) for stochastic processes in general among others. The problem of interest is to study the rate of convergence of the maximum likelihood estimator (MLE) $\hat{\theta}_n$ of the parameter θ based on the observation $(X_1, \ldots, X_n)$. Results on moderate deviations for the maximum likelihood estimator for the case of independent and identically distributed observations were proved by Gao (2001) and for the case of independent but possibly not identically distributed observations by Xiao and Liu (2006). Miao and Chen (2010) gave a simpler proof to obtain these results under weaker conditions using Gärtner-Ellis theorem (cf. Hollander (2000), Theorem V.6). Miao and Wang (2014) improved the result in Miao and Chen (2010) by weakening the exponential integrability condition.

Our aim in this paper is to extend the results in Miao and Chen (2010) to maximum likelihood estimator for Markov processes. We give a short introduction to maximum likelihood estimation for Markov processes due to Billingsley (1961) for completeness and to introduce the notation.

Suppose the process $\{X_n, n \geq 1\}$ is a Markov process for each $\theta \in \Theta \subset R$, with stationary transition measure

(1. 1) $$p_\theta(x, A) = P_\theta(X_{n+1} \in A | X_n = x), A \in \mathcal{F}_X.$$

We assume that, for each $\theta \in \Theta$, the function $p_\theta(x, A)$ is a measurable function of x for each fixed $A \in \mathcal{F}_X$ and a probability measure on $\mathcal{F}_X$ for fixed x. It is known that such a set of transition measures give rise to a Markov process with stationary transition measure given by (1.1)(cf. Doob (1953)). We assume that, for each $\theta \in \Theta$, the transition measures admit a unique stationary probability distribution, that is, there is a unique probability measure $p_\theta(.)$ on $\mathcal{F}_X$ such that

$$p_\theta(A) = \int_{\mathcal{X}} p_\theta(x, A)\, p_\theta(dx), A \in \mathcal{F}_X.$$

Here after $E_\theta(.)$ will denote the expectation computed under the assumption that θ is the true parameter. We will not assume that $p_\theta(.)$ is the initial distribution. The initial distribution has no effect on the conditional expectation $E_\theta(.|X_1)$ as the conditional expectation involves only the transition probability measure. We will assume that there is a σ-finite measure λ on $(\mathcal{X}, \mathcal{F}_X)$ with respect to which all the transition measures have densities $f(x, y; \theta)$. Hence

$$p_\theta(x, A) = \int_A f(x, y; \theta)\, \lambda(dy),\ A \in \mathcal{F}_X.$$

We will assume that the initial distribution has a density $f(x; \theta)$ with respect to λ. We assume that the function $f(x; \theta)$ is jointly measurable in (x, θ) and the function $f(x, y; \theta)$ is jointly measurable in (x, y, θ).

Suppose $(x_1, \ldots, x_n)$ is an observation on the discrete time Markov process observed up to time n. Then the log-likelihood function of the observation $(x_1, \ldots, x_n)$ is

$$\log f(x_1; \theta) + \sum_{k=1}^{n-1} \log f(x_k, x_{k+1}; \theta).$$

The term $\log f(x_1; \theta)$ in the likelihood function is dominated by the other terms in the log-likelihood function as n tend to infinity and the information about the parameter θ in the initial observation can be ignored as we are studying the large sample properties of the estimators for the parameter θ. Hence, we will take the log-likelihood function, here after, to be

$$\ell_n(x_1, \ldots, x_n; \theta) = \sum_{k=1}^{n-1} \log f(x_k, x_{k+1}; \theta).$$

If we assume that the initial observation x_1 is a constant and does not depend on the parameter θ, then the above expression will be the exact log-likelihood. Suppose the following regularity conditions hold:

(C0) The parameter space Θ is open in R.

(C1) For any x, the set of y for which $f(x, y; \theta) > 0$ does not depend on the parameter θ.

(C2) For any x and y, the function $f(x, y; \theta)$ is thrice differentiable for $\theta \in \Theta$ and the derivatives are continuous in $\theta \in \Theta$. Here after we denote the i-th derivative of $f(x, y; \theta)$ with respect to θ evaluated at θ' as $f^{(i)}(x, y; \theta')$ and let $\ell(x, y; \theta) = \log f(x, y; \theta)$.

(C3) For any $\theta \in \Theta$, there exists a neighbourhood $G(\theta, \delta)$ of θ for some $\delta > 0$, such that

$$\int_{\mathcal{X}} \sup_{\theta' \in G(\theta,\delta)} |f^{(i)}(x, y; \theta')| \lambda(dy) < \infty, i = 1, 2,$$

and

$$E_\theta[\sup_{\theta' \in G(\theta,\delta)} |\ell^{(3)}(X_1, X_2; \theta')|] < \infty.$$

(C4) Furthermore

$$0 \leq E_\theta[|\ell^{(1)}(X_1, X_2; \theta)|^2] < \infty.$$

Let $I(X_k; \theta)$ denote the conditional Fisher information in the observation in X_{k+1} given the observations $X_i, 1 \leq i \leq k$ or equivalently X_k by the Markov property of the process $\{X_i, i \geq 1\}$ when the true parameter is θ.

In view of Theorem 1.1 stated below, it follows that

$$\frac{1}{n} \sum_{k=1}^{n-1} I(X_k; \theta)$$

tends to a limit, say, $I(\theta)$ a.s. as $n \to \infty$. This limit does not depend on the initial distribution of the Markov process. Suppose that $0 < I(\theta) < \infty$.

In addition to the conditions (C1) to (C4), we assume the following condition holds:

(C5) For each $\theta \in \Theta$, the stationary distribution $p_\theta(.)$ exists and is unique and has the property, that for each $x \in \mathcal{X}$, the probability measure corresponding to the probability density function $p_\theta(x,.)$ is absolutely continuous with respect to the probability measure corresponding to the probability density function $p_\theta(.)$.

Billingsley (1961) proved the following strong law of large numbers for Markov processes.

Theorem 1.1: *Suppose the condition (C5) holds. Then, no matter what the initial distributions is, if $\phi(x,y)$ is measurable with respect to $\mathcal{F}_X \times \mathcal{F}_X$, and if $E_\theta(|\phi(X_1,X_2)|) < \infty$, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n-1} \phi(X_k, X_{k+1}) = E_\theta(\phi(X_1,X_2))$$

almost surely.

In view of the conditions (C1) to (C5), it follows that

$$E_\theta(\ell^{(1)}(X_k, X_{k+1};\theta)|X_k) = 0$$

almost surely and the partial sums

$$\sum_{k=1}^{n-1} \ell^{(1)}(X_k, X_{k+1};\theta), n \geq 2$$

form a martingale under the probability measure P_θ. An application of central limit theorem for martingales due to Billingsley (1961) or Ibragimov (1963) will show that the sequence

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} \ell^{(1)}(X_k, X_{k+1};\theta)$$

is asymptotically normal with mean zero and variance $I(\theta) < \infty$. The following theorem is due to Billingsley (1961).

Theorem 1.2: *Suppose the conditions (C1)-(C5) hold. Then there exists a sequence $\hat{\theta}_n$ of random variables depending on the observations $(x_1,\ldots,x_n)$ such that $\hat{\theta}_n$ converges in probability to the true parameter θ and such that $\hat{\theta}_n$ is a solution of the likelihood equation*

$$\frac{d}{d\theta}\ell_n(x_1,\ldots,x_n;\theta) = \sum_{k=1}^{n-1} \ell^{(1)}(x_k, x_{k+1};\theta) = 0.$$

Furthermore

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \frac{1}{I(\theta)}) \ \ as \ \ n \to \infty.$$

Here $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. For alternate sufficient conditions for the asymptotic normality of the maximum likelihood estimator for discrete time Markov processes, see Prakasa Rao (1972). Let $\epsilon > 0$. We now study the rate of convergence of the probability $P_\theta(\lambda(n)|\hat{\theta}_n - \theta| \geq \epsilon)$ where $\lambda(n) \to \infty$ and $\frac{\lambda(n)}{\sqrt{n}} \to 0$ as $n \to \infty$, the problem of moderate deviation.

2 Main results

Let Θ be an open interval in R and the process $\{X_k, k \geq 1\}$ be a discrete time homogeneous Markov process with the transition density $f(x, y; \theta), \theta \in \Theta$ which is continuously differentiable with respect to θ. Let $\ell(x, y; \theta) = \log f(x, y; \theta)$. Define

$$\ell_n(x_1, \dots, x_n; \theta) = \log \prod_{k=1}^{n-1} f(x_k, x_{k+1}; \theta) = \sum_{k=1}^{n-1} \ell(x_k, x_{k+1}; \theta)$$

and, for any $n \geq 1$,

$$\ell_n^{(1)}(x_1, \dots, x_n; \theta) = \frac{\partial \ell_n(x_1, \dots, x_n; \theta)}{\partial \theta}, (x_1, \dots, x_n) \in R^n.$$

We define the maximum likelihood estimator $\hat{\theta}_n(x_1, \dots, x_n)$ as a solution of the equation

$$\ell_n^{(1)}(x_1, \dots, x_n; \theta) = 0.$$

Let

$$\underline{\theta}_n = \underline{\theta}_n(x_1, \dots, x_n) = \inf\{\theta \in \Theta : \ell_n^{(1)}(x_1, \dots, x_n; \theta) \leq 0\}$$

and

$$\bar{\theta}_n = \bar{\theta}_n(x_1, \dots, x_n) = \sup\{\theta \in \Theta : \ell_n^{(1)}(x_1, \dots, x_n; \theta) \geq 0\}.$$

It is obvious that

$$\underline{\theta}_n(x_1, \dots, x_n) \leq \hat{\theta}_n(x_1, \dots, x_n) \leq \bar{\theta}_n(x_1, \dots, x_n)$$

and, for every $\epsilon > 0$,

$$P_\theta(\underline{\theta}_n \geq \theta + \epsilon) \leq P_\theta(\ell_n^{(1)}(x_1, \dots, x_n; \theta + \epsilon) \geq 0) \leq P_\theta(\bar{\theta}_n \geq \theta + \epsilon)$$

and

$$P_\theta(\bar{\theta}_n \leq \theta - \epsilon) \leq P_\theta(\ell_n^{(1)}(x_1, \dots, x_n; \theta - \epsilon) \leq 0) \leq P_\theta(\underline{\theta}_n \leq \theta - \epsilon).$$

We assume that the following conditions hold:

$(C1')$ For each $\theta \in \Theta$, the derivatives

$$\ell^{(i)}(x, y; \theta) = \frac{\partial^i \log f(x, y; \theta)}{\partial \theta^i}, i = 1, 2, 3$$

exists for all $x, y \in R$.

$(C2')$ For each $\theta \in \Theta$, there exists a neighbourhood $G(\theta, \delta)$ of θ for some $\delta > 0$ and non-negative measurable functions $A_i(x, y; \theta), i = 1, 2, 3$ such that

$$\sup_{x \in R} \int_R [A(x, y; \theta)]^6 \, f(x, y; \theta) dy < \infty, i = 1, 2, 3$$

and

$$\sup_{\theta' \in G(\theta, \delta)} |\ell^{(i)}(x, y; \theta')| \leq A(x, y; \theta), i = 1, 2, 3.$$

$(C3')$ For each $\theta \in \Theta$, the transition probability density function $f(x, y; \theta)$ has a finite non-negative conditional Fisher information, that is,

$$0 \leq I(x; \theta) = E_\theta[(\frac{\partial \log f(X_1, X_2; \theta)}{\partial \theta})^2 | X_1 = x] < \infty$$

for all $x \in R$.

$(C4')$ For each $\theta \in \Theta$, there exists a function $0 < I(\theta) < \infty$ such that

$$\frac{1}{n} \sum_{k=1}^{n-1} E_\theta(I(X_k; \theta)) \to I(\theta)$$

as $n \to \infty$.

$(C5')$ For each $\theta \in \Theta$, there exists $\mu = \mu(\theta)$ and $\nu = \nu(\theta)$ such that

$$\sup_{(t,\epsilon) \in [-\mu,\mu] \times [-\nu,\nu]} \phi(t; \theta, \epsilon) < \infty,$$

where

$$\phi(t; \theta, \epsilon) = \sup_x E_\theta[\exp(t\ell^{(1)}(X_1, X_2; \theta + \epsilon)) | X_1 = x].$$

$(C6')$ For all $(x_1, \ldots, x_n) \in R^n, n \geq 1$, the likelihood equation

$$\ell_n^{(1)}(x_1, \ldots, x_n; \theta) = 0$$

has a unique solution.

Under the conditions $(C1') - (C3')$, it can be checked that,

$$E_\theta(\ell^{(1)}(X_1, X_2; \theta)|X_1 = x) = 0 \quad \text{a.s.} \tag{2. 1}$$

and

$$E_\theta[(\ell^{(1)}(X_1, X_2; \theta))^2|X_1 = x] = -E_\theta[\ell^{(2)}(X_1, X_2; \theta)|X_1 = x] = I(x; \theta) \quad \text{a.s.} \tag{2. 2}$$

$(C7')$ Suppose that $\{\lambda(n), n \geq 1\}$ is a sequence of positive numbers such that $\lambda(n) \to \infty$, $\frac{\lambda(n)}{\sqrt{n}} \to 0$ as $n \to \infty$ and

$$E_\theta[|\sum_{k=1}^{n-1} \ell^{(1)}(X_k, X_{k+1}; \theta + \frac{\epsilon}{\lambda(n)})|^3] = o(n\lambda(n)).$$

Theorem 2.1: *Under the conditions $(C1')$ to $(C7')$,*

$$\liminf_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)(\bar{\theta}_n - \theta) \geq \epsilon) \geq -\frac{1}{2}I(\theta)\epsilon^2, \tag{2. 3}$$

$$\liminf_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)(\underline{\theta}_n - \theta) \leq -\epsilon) \geq -\frac{1}{2}I(\theta)\epsilon^2, \tag{2. 4}$$

$$\limsup_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)(\underline{\theta}_n - \theta) \geq \epsilon) \leq -\frac{1}{2}I(\theta)\epsilon^2, \tag{2. 5}$$

and

$$\limsup_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)(\bar{\theta}_n - \theta) \leq -\epsilon) \leq -\frac{1}{2}I(\theta)\epsilon^2. \tag{2. 6}$$

Furthermore

$$\lim_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)|\hat{\theta}_n - \theta| \geq \epsilon) = -\frac{1}{2}I(\theta)\epsilon^2. \tag{2. 7}$$

The following theorem is a consequence of Theorem 2.1.

Theorem 2.2: *Suppose that the conditions $(C1')$ to $(C6')$ hold. Then, for any closed subset $F \subset \Theta$,*

$$\limsup_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)(\hat{\theta}_n - \theta) \in F) \leq -\frac{1}{2}I(\theta) \inf_{x\in F} x^2, \tag{2. 8}$$

and, for any open subset $G \subset \Theta$,

$$\liminf_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)(\hat{\theta}_n - \theta) \in G) \geq -\frac{1}{2}I(\theta) \inf_{x\in G} x^2, \tag{2. 9}$$

and, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)|\hat{\theta}_n - \theta| \geq \epsilon) = -\frac{1}{2}I(\theta)\epsilon^2, \tag{2. 10}$$

3 Proofs of Theorem 2.1 and Theorem 2.2

We now prove the following lemma.

Lemma 3.1: *Under the conditions* $(C1')$ *to* $(C5')$, *for any* $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\ell_n^{(1)}(X_1,\ldots,X_n;\theta + \frac{\epsilon}{\lambda(n)}) \geq 0) = -\frac{I(\theta)\epsilon^2}{2} \tag{3. 1}$$

and

$$\lim_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\ell_n^{(1)}(X_1,\ldots,X_n;\theta - \frac{\epsilon}{\lambda(n)}) \leq 0) = -\frac{I(\theta)\epsilon^2}{2}. \tag{3. 2}$$

Proof: Applying the Taylor's expansion of $\ell^{(1)}(x,y;\theta)$ in the neighbourhood $N(\theta,\delta)$, we get that

$$\sup_{x\in R} |\ell^{(1)}(x,y;\gamma) - \ell^{(1)}(x,y;\theta) - (\gamma-\theta)\ell^{(2)}(x,y;\theta)| \leq \frac{1}{2}(\gamma-\theta)^2 A_3(x,y;\theta).$$

Hence, for any $k \geq 1$,

$$|\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}) - \ell^{(1)}(X_k,X_{k+1};\theta) - \frac{\epsilon}{\lambda(n)}\ell^{(2)}(X_k,X_{k+1};\theta)| \leq \frac{\epsilon^2}{2\lambda^2(n)} A_3(X_k,X_{k+1};\theta).$$

Hence, by the condition $(C2')$ and the equations (2.1) and (2.2), it follows that

(3. 3)

$$\begin{aligned} E_\theta[\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})] &= E_\theta[\ell^{(1)}(X_k,X_{k+1};\theta)] + \frac{\epsilon}{\lambda(n)} E_\theta[\ell^{(2)}(X_k,X_{k+1};\theta)] + o(\frac{1}{\lambda(n)}) \\ &= -E_\theta[I(X_k;\theta)]\frac{\epsilon}{\lambda(n)} + o(\frac{1}{\lambda(n)}). \end{aligned}$$

Therefore, it follows that

$$\begin{aligned} & P_\theta(\ell_n^{(1)}(X_1,\ldots,X_n;\theta+\frac{\epsilon}{\lambda(n)}) \geq 0) \\ = \;& P_\theta[\frac{\lambda(n)}{n}\sum_{k=1}^{n-1}(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}) - E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})) \\ & \quad \geq -\frac{\lambda(n)}{n}\sum_{k=1}^{n-1} E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}))] \\ = \;& P_\theta[\frac{\lambda(n)}{n}\sum_{k=1}^{n-1}(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}) - E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})) \\ & \quad \geq \frac{\sum_{k=1}^{n-1} E_\theta(I(X_k,\theta))}{n-1}\epsilon + o(1)] \;\; \text{(by (3.3))} \end{aligned}$$

We now compute the functional

$$\lim_{n\to\infty}\frac{\lambda^2(n)}{n}\log E_\theta\{\exp(\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}[\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})-E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}))])\}$$

for any $t\in R$. Applying the inequality

$$|e^x-1-x-\frac{x^2}{2}|\le |x|^3e^{|x|}, x\in R$$

and the condition $(C5')$, it follows that, for every $k\ge 1$, and for $t\in R$,

(3. 4)

$$\begin{aligned}E_\theta\{\exp(\frac{t}{\lambda(n)}\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}))\} &= 1+\frac{t}{\lambda(n)}E_\theta[\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})]\\ &+\frac{t^2}{2\lambda^2(n)}E_\theta([\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})]^2)+o(\frac{1}{\lambda^2(n)}).\end{aligned}$$

Let

$$J_{n,\epsilon}\equiv\sum_{k=1}^{n-1}\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}),$$

$$H_n\equiv\sup_{\theta'\in G(\theta,\delta)}|\sum_{k=1}^{n-1}\ell^{(2)}(X_k,X_{k+1};\theta')|,$$

and

$$K_n\equiv\sup_{\theta'\in G(\theta,\delta)}|\sum_{k=1}^{n-1}\ell^{(3)}(X_k,X_{k+1};\theta')|.$$

Observe that

$$|J_{n,\epsilon}|\le J_n\equiv\sum_{k=1}^{n-1}A(X_k,X_{k+1};\theta)=O_p(n),$$

$$H_n\le J_n=O_p(n)$$

and

$$K_n\le J_n=O_p(n)$$

by the condition $(C2')$ and Theorem 1.1. Note that, for every $\epsilon>0$,

$$J_{n,\epsilon}=\sum_{k=1}^{n-1}\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})=o_p((n\lambda(n))^{1/3})$$

by the condition $(C7')$. Then, by the conditions $(C2')$ and $(C5')$ and the equations (2.1) and (2.2), it follows that

(3. 5)

$$
\begin{aligned}
&\log E_\theta(\exp(\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}[\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})-E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}))]))\\
=\ & \log[E_\theta\{\exp(\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)}))\}]\\
&+(-\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})))\\
=\ & \log E_\theta[1+\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})\\
&+\frac{t^2}{2\lambda^2(n)}[\sum_{k=1}^{n-1}\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})]^2\\
&+O_p(\frac{t^3}{\lambda^3(n)}|J_{n\epsilon}|^3)] \qquad (3.\ 6)\\
&+(-\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})))\\
=\ & \log E_\theta[1+\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}\ell^{(1)}(X_k,X_{k+1};\theta)\\
&+\frac{t\epsilon}{\lambda^2(n)}\sum_{k=1}^{n-1}\ell^{(2)}(X_k,X_{k+1};\theta)+O_p(\frac{t\epsilon^2}{\lambda^3(n)}K_n)\\
&+\frac{t^2}{2\lambda^2(n)}[\sum_{k=1}^{n-1}\ell^{(1)}(X_k,X_{k+1};\theta)]^2+O_p(\frac{t^2\epsilon}{\lambda^3(n)}J_n^{1/2}H_n^{1/2})]+O_p(\frac{t^3}{\lambda^3(n)}|J_{n\epsilon}|^3)\\
&+(-\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})))\\
=\ & \log[1+(\frac{-t\epsilon}{\lambda^2(n)}+\frac{t^2}{2\lambda^2(n)})\sum_{k=1}^{n-1}E_\theta(I(X_k;\theta))+O(\frac{t^3\epsilon}{\lambda^3(n)}o_p(n\lambda(n)))]\\
&+(-\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})))\\
=\ & (\frac{-t\epsilon}{\lambda^2(n)}+\frac{t^2}{2\lambda^2(n)})\sum_{k=1}^{n-1}E_\theta(I(X_k;\theta))+O(\frac{t^3\epsilon}{\lambda^3(n)}o_p(n\lambda(n)))\\
&+(-\frac{t}{\lambda(n)}\sum_{k=1}^{n-1}E_\theta(\ell^{(1)}(X_k,X_{k+1};\theta+\frac{\epsilon}{\lambda(n)})))
\end{aligned}
$$

$$
\begin{aligned}
&= (\frac{-t\epsilon}{\lambda^2(n)} + \frac{t^2}{2\lambda^2(n)}) \sum_{k=1}^{n-1} E_\theta(I(X_k;\theta)) + O(\frac{t^3\epsilon}{\lambda^3(n)} o_p(n\lambda(n)))) \\
&\quad +(\frac{t\epsilon}{\lambda^2(n)} \sum_{k=1}^{n-1} E_\theta(I(X_k,\theta)) + O(\frac{t\epsilon^2}{\lambda^3(n)} K_n) \\
&= \frac{t^2(n-1)}{2\lambda^2(n)} I(\theta) + O(\frac{t^3\epsilon}{\lambda^3(n)} o_p(n\lambda(n)))).
\end{aligned}
$$

which implies that

$$
\lim_{n\to\infty} \frac{\lambda^2(n)}{n} \log E_\theta\{\exp(\frac{t}{\lambda(n)} \sum_{k=1}^{n-1} [\ell^{(1)}(X_k, X_{k+1}; \theta + \frac{\epsilon}{\lambda(n)}) - E_\theta(\ell^{(1)}(X_k, X_{k+1}; \theta + \frac{\epsilon}{\lambda(n)})]])\}
$$

$$
= \frac{I(\theta)}{2} t^2.
$$

Applying the Gärtner-Ellis theorem (cf. Gärtner (1977); Ellis (1984); Hollander (2000), Theorem V.6), we get the result stated in equation (3.1) of Lemma 3.1. Similar analysis will prove equation (3.2).

Proof of Theorem 2.1: Observe that

$$
P_\theta(\lambda(n)(\bar{\theta}_n - \theta) \geq \epsilon) \geq P_\theta(\ell_n^{(1)}(X_1, \ldots, X_n; \theta + \frac{\epsilon}{\lambda(n)} \geq 0)
$$

and

$$
P_\theta(\lambda(n)(\underline{\theta}_n - \theta) \geq \epsilon) \leq P_\theta(\ell_n^{(1)}(X_1, \ldots, X_n; \theta + \frac{\epsilon}{\lambda(n)} \geq 0).
$$

An application of Lemma 3.1 implies the relations (2.3) and (2.5). Similar arguments prove (2.4), (2.6) and (2.7).

Proof of Theorem 2.2 is analogous to the proof of a similar result in Gao (2001) in the work on moderate deviation of the maximum likelihood estimator in the independent and identically distributed case.

Proof of Theorem 2.2 : For any closed subset $F \subset \Theta \subset R$, define $x_1 = \inf\{x > 0 : x \in F\}$ and $x_2 = \sup\{x < 0 : x \in F\}$. Let $I(y;\theta) = \frac{1}{2} I(\theta) y^2$.Then

$$
\limsup_{n\to\infty} \frac{\lambda^2(n)}{n} \log P_\theta(\lambda(n)(\hat{\theta}_n - \theta) \in F)
$$

$$\leq \limsup_{n\to\infty} \frac{\lambda^2(n)}{n} \log(P_\theta(\lambda(n)(\hat{\theta}_n - \theta) \leq x_2) + P_\theta(\lambda(n)(\hat{\theta}_n - \theta) \geq x_1))$$
$$\leq \max(-I(x_2;\theta), -I(x_1;\theta)) = -\inf_{x\in F} I(x;\theta) = -\frac{1}{2} I(\theta) \inf_{x\in F} x^2.$$

Suppose G is an open subset of $\Theta \subset R$. Then, for any $\epsilon > 0$ such that $(x-\epsilon, x+\epsilon) \subset G$, it follows that

$$\begin{aligned} & P_\theta(\lambda(n)(\hat{\theta}_n - \theta) \in G) \\ \geq\ & P_\theta(x - \epsilon \leq \lambda(n)(\hat{\theta}_n - \theta) \leq x + \epsilon) \\ =\ & P_\theta[\sum_{k=1}^n \ell^{(1)}(X_k, X_{k+1}; \theta + \frac{x+\epsilon}{\lambda(n)} \leq 0, \sum_{k=1}^n \ell^{(1)}(X_k, X_{k+1}; \theta + \frac{x-\epsilon}{\lambda(n)} \geq 0]. \end{aligned}$$

Note that the sequence $\{S_n = \sum_{k=1}^n \ell^{(1)}(X_k, X_{k+1}; \theta), n \geq 1\}$ is a zero mean martingale under P_θ-measure. Applying martingale central limit theorem, it follows that

$$P_\theta(\frac{1}{\sqrt{n}} \sum_{k=1}^n \ell^{(1)}(X_k, X_{k+1}; \theta) \geq \eta t) \to 1 - \Phi(\frac{\eta t}{\sigma_\theta})$$

for some $\sigma_\theta > 0$. By arguments similar to those given in the proof of Theorem 2.1, it follows that, for all $t > 0$ and $\eta > 0$,

$$\liminf_{n\to\infty} P_\theta(\lambda(n)(\hat{\theta}_n - \theta) \in G)$$
$$\geq \frac{1}{t^2} \log(\Phi(t(\eta + I(\theta)(x+\epsilon))/\sigma_\theta) - \Phi(t(\eta + I(\theta)(x-\epsilon))/\sigma_\theta)).$$

Let $\eta \to 0$ at first, then let $t \to \infty$ and let $\epsilon \to 0$. Then we get that

$$\liminf_{n\to\infty} P(_\theta(\lambda(n)(\hat{\theta}_n - \theta) \in G) \geq -\frac{1}{2} I(\theta) x^2. \tag{3. 7}$$

Since $x \in G$ is arbitrary in the above discussion, it follows that

$$\liminf_{n\to\infty} P(_\theta(\lambda(n)(\hat{\theta}_n - \theta) \in G) \geq -\frac{1}{2} I(\theta) \inf_{x\in G} x^2. \tag{3. 8}$$

We will present an example to illustrate the results.

Example: Let $\{X_n, n \geq 1\}$ be a stochastic process defined recursively by the relation

$$X_{n+1} = \theta X_n + Y_{n+1}$$

where $\{X_1, Y_2, Y_3, \ldots\}$ is an independent sequence of random variables. Further suppose that $|\theta| < 1$ and the random variable Y_n is standard normal for every $n \geq 2$. This is the first-order linear autoregressive model and it is known that there exists a unique stationary distribution for the process and it is normal with mean zero and variance $\frac{1}{1-\theta^2}$. Suppose that the random variable X_1 has this stationary distribution. Then the process $\{X_n, n \geq 1\}$ is a stationary Markov process. It can be checked that the logarithm of the transition density function $f(x_k, x_{k+1}; \theta)$ is given by

$$\ell(x_k, x_{k+1}; \theta) = \log(2\pi)^{-1/2} - \frac{1}{2}(x_{k+1} - \theta x_k)^2.$$

It is easy to check that the first three derivatives of $\ell(x_k, x_{k+1}; \theta)$ exist and are given by

$$\ell^{(1)}(x_k, x_{k+1}; \theta) = x_k(x_{k+1} - \theta x_k),$$

$$\ell^{((2)}(x_k, x_{k+1}; \theta) = -x_k^2,$$

and

$$\ell^{(3)}(x_k, x_{k+1}; \theta) = 0.$$

It can now be seen the the conditions $(C1')$ to $(C6')$ hold in this example from the properties of the Gaussian distribution and hence the results stated in Theorem 2.2 hold for the maximum likelihood estimator for θ with the function $I(\theta) = \frac{1}{1-\theta^2}$.

Acknowledgement: The author thanks the referee for pointing out a lacuna in an earlier version of this paper. This work was supported under the scheme "Ramanujan Chair Professor" at the CR Rao Advanced Institute for Mathematics, Statistics and Computer science, Hyderabad, India.

References:

Basawa, I.V. and Prakasa Rao, B.L.S. (1980) *Statistical Inference for Stochastic Processes*, Academic Press, London.

Billingsley, P. (1961) *Statistical Inference for Markov Processes*, The University of Chicago Press, Chicago.

Doob, J.L. (1953) *Stochastic Processes*, Wiley, New York.

Ellis, R.S. (1984) Large deviations for a general class of random vectors, *Ann. Probab.*, **12**, 1-12.

Gärtner, J. (1977) On large deviations from the invariant measure, *Theory Probab. Appl.*, **22**, 24-39.

Gao, Fuqing. (2001) Moderate deviations for the maximum likelihood estimator, *Statist. Probab. Lett.*, **55**, 345-352.

Grenander, U. (1981) *Abstract Inference*, Wiley, New York.

Hollander, F. den (2000) *Large Deviations*, Field Institute Monographs, American Mathematical Society, Providence, Rhode Island.

Ibragimov, I.A. (1963) A central limit theorem for a class of dependent random variables, *Theory. Probab. Appl.*, **8**, 83-89.

Miao, Y. and Chen, Ying-Xia (2010) Note on the moderate deviation principle of maximum likelihood estimator, *Acta Appl. Math.*, **110**, 863-869.

Miao, Yu and Wang, Yanling (2014) Moderate deviation principle for maximum likelihood estimator, *Statistics*, **48**, 766-777.

Prakasa Rao, B.L.S. (1972) Maximum likelihood estimation for Markov processes, *Ann. Inst. Statist. Math.*, **24**, 333-345.

Prakasa Rao, B.L.S. (1973) On the rate of convergence of estimators for Markov processes, *Z. Wahr. Verw Gebiete*, **26**, 141-152.

Prakasa Rao, B.L.S. (1979) The equivalence between (modified) Bayes and maximum likelihood estimators for Markov processes, *Ann. Inst. Statist. Math.*, **31**, 499-513.

Xiao, Zhihong and Liu, Luqin. (2006) Moderate deviations of maximum likelihood estimator for independent not identically distributed case, *Statist. Probab. Lett.*, **76**, 1056-1064.