Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

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

ARTICLE INFO

Article history:
Received 22 March 2017
Received in revised form 7 September 2017
Accepted 14 September 2017
Available online xxxx

MSC 2010:
62M05
62F12
60F99

Keywords:
Moderate deviation
Maximum likelihood estimation
Markov process

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.

1. Introduction

Let $\{X_n, n \geq 1\}$ be a stochastic process defined on a probability 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, 1979) 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

$$p_\theta(x, A) = P_\theta(X_{n+1} \in A | X_n = x), A \in \mathcal{F}_X. \tag{1.1}$$

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

E-mail address: blsprao@gmail.com.

http://dx.doi.org/10.1016/j.spl.2017.09.009
0167-7152/

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 tends 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 ith 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.