Journal of Statistical Planning and Inference 139 (2009) 2242 – 2253

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Variance estimation in the central limit theorem for Markov chains

Samis Trevezas*, Nikolaos Limnios

Laboratoire de Mathématiques Appliquées de Compiègne, Université de Technologie de Compiègne, France

ARTICLE INFO

Article history:
Received 8 January 2007
Received in revised form
12 October 2007
Accepted 17 October 2008
Available online 7 November 2008

MSC:
62M05
60F05
60J10
62F12

Keywords:
Markov chain estimation
Central limit theorem
Variance estimation

ABSTRACT

This article concerns the variance estimation in the central limit theorem for finite recurrent Markov chains. The associated variance is calculated in terms of the transition matrix of the Markov chain. We prove the equivalence of different matrix forms representing this variance. The maximum likelihood estimator for this variance is constructed and it is proved that it is strongly consistent and asymptotically normal. The main part of our analysis consists in presenting closed matrix forms for this new variance. Additionally, we prove the asymptotic equivalence between the empirical and the maximum likelihood estimation (MLE) for the stationary distribution.

1. Introduction and preliminaries

Markov processes (MP) are a central topic in applied probability and statistics. The reason is that many real problems can be modeled by this kind of stochastic processes in continuous or in discrete time. Statistical estimation for finite state ergodic Markov chains (MC) was discussed by Billingsley (1961a,b). Estimation consists in considering a trajectory observed on a time interval $\{0, 1, \ldots, m\}$ and then give an estimate for the transition function. Other quantities can also be estimated, as the stationary distribution, the mean hitting times and relevant probabilistic characteristics which are involved in the analysis of an MC. To this purpose, an important problem is to estimate the variance σ^2 which appears in CLTs. In particular, when we consider additive functionals of an MC, this variance is the same as in the functional central limit theorems (FCLT) for MC. Recent papers concentrate on this aspect for MCMC settings. A recent survey is due to Roberts and Rosenthal (2004). Chauveau and Diebolt (2003) study a continuous-time empirical variance process based on i.i.d. parallel chains. Stefanov (1995) considers also maximum likelihood estimation (MLE) for MP and MC via the exponential families approach and derives closed form solutions for the variance parameters in FCLTs. The CLT is still an aspect of the MC which worth being studied even in a countable or in a finite state space. To this end our work focus the attention on computational aspects for the MLE of the associated variance. This paper is organized as follows. In Section 2 we study and present the associated variance on the CLT in different matrix forms. These different forms result from different proofs of the CLT found already in the literature with some slight variations in the terminology. For some different proofs see Doob (1953), Billingsley (1961b), Dacunha-Castelle and Duflo (1983), Meyn and Tweedy (1993) and Port (1994). The starting point for our analysis is the form for σ^2 given in Port (1994). In the first part of Section 3 we review some basic facts on the estimation of the stationary distribution and we prove the equivalence between the MLE and the empirical

* Corresponding author. Tel.: +33 3 44 23 44 23; fax: +33 3 44 23 44 77.
E-mail address: streveza@utc.fr (S. Trevezas).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.10.020

estimator of the stationary distribution. Our main task is accomplished in the second part of Section 3, where the MLE of the variance is constructed in one of its matrix forms and proved to be strongly consistent and asymptotically normal with asymptotic variance computed in two different matrix forms. Finally, we present some examples in order to enlighten the theoretical results.

Let $\mathbf{X}=(X_n;n\in\mathbb{N})$ be an MC which we assume (except for Theorem 1 which is more generally valid on countable state space) to have finite state space $E=\{1,2,\dots,s\}$ and be defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$. The values belonging to E will represent here the states of an observed system. It is assumed that X is irreducible with transition matrix P and stationary distribution $\boldsymbol{\pi}$. We are going to use the following notation:

- $f:E\to\mathbb{R}$; f will be understood here as a reward function associated with the states of the system.
- $S_n=\sum_{i=0}^{n}f(X_i)$; S_n stands for the $n+1$ partial sum of the reward process $(f(X_n);\ n\in\mathbb{N})$.
- $\mu=\sum_{i\in E}f(i)\pi(i)=\mathbb{E}_\pi f(X_0)$; μ indicates the expectation of $f(X_0)$ with respect to the stationary distribution of $\mathbf{X}$.
- $\widetilde{f}(i)=f(i)-\mu$; $\widetilde{f}$ will denote f centralized with respect to the stationary distribution of $\mathbf{X}$.
- $\boldsymbol{e}_i$ are the vectors from the orthonormal base of $\mathbb{R}^s$.
- $T_a=\inf\{n>0:X_n=a\}$, $a\in E$; T_a records the first hitting time of state a disregarding the initial state.
- $\mathbb{P}_i(A)=\mathbb{P}(A|X_0=i)$ where $A\in\mathscr{F}$ (i.e., A represents a measurable event), $\mathbb{E}_i$ the corresponding expectation operator, $i\in E$.
- $g_a(i,j)=\sum_{n=1}^{+\infty}\mathbb{P}_i(T_a\geqslant n,X_n=j)$.
- P_a is a $(s-1)\times(s-1)$ matrix which results from P excluding both its a-column and a-row.
- P'_a is a $s\times(s-1)$ matrix which results from P excluding only its a-column.
- $\mathbf{1}$ is an s-dimensional row vector, having all his elements equal to 1.
- $A=\mathbf{1}^\top\mathbf{1}$ and I denotes the s-dimensional identity matrix.

Note that

$$g_a(i,j)=\sum_{n=1}^{+\infty}\mathbb{E}(1\{T_a\geqslant n,X_n=j\}|X_0=i)=\mathbb{E}\left(\sum_{n=1}^{+\infty}1\{T_a\geqslant n,X_n=j\}|X_0=i\right)=\mathbb{E}_i(N_j^a),\tag{1}$$

where N_j^a is an r.v. which counts, after the system leaves from the initial state, how many times it has visited state j, until it reaches for the first time state a.

The above expression of $g_a(i,j)$ gives a concrete probabilistic meaning to this quantity and should be distinguished from, $g_a^*(i,j)=\sum_{n=0}^{+\infty}\mathbb{P}_i(T_a\geqslant n,X_n=j)$, which is often used in the literature and will be mentioned also later in this paper. It is obvious that for $a,i,j\in E$,

$$g_a^*(i,j)=\delta_{ij}+g_a(i,j).\tag{2}$$

The function $g_a(i,j)$ will play a fundamental role in the analysis of the variance on the CLT. Equivalent representations of this variance already exist, and we will indicate some of them in matrix form. These forms can be expressed as a function of the transition probability matrix P of the MC together with the reward function f. Therefore, all these forms can serve as plug-in type estimators for the variance. For the problem of estimation we will choose one form involving the notion of hitting times. By proving the asymptotic normality we gain different representations of the asymptotic variance of these estimators. The following theorem is the starting point of our work.

Theorem 1 (Port, 1994). *Assume that for some state $a\in E$, $\sum_{i,j\in E}\pi(i)g_a(i,j)|\widetilde{f}(i)||\widetilde{f}(j)|<\infty$ and set*

$$\sigma^2=\sum_{i\in E}\pi(i)\widetilde{f}(i)^2+2\sum_{\substack{i\neq a\\ i,j\in E}}\pi(i)g_a(i,j)\widetilde{f}(i)\widetilde{f}(j).\tag{3}$$

Then for any initial distribution, $(1/\sqrt{n})(S_n-n\mu)\xrightarrow{d}\sigma Z$ where Z is standard normally distributed. Here $E=\mathbb{N}$.

For a proof of this theorem see Port (1994).

Remark 1. It is also proved easily (see Port, 1994) that the validity of the condition stated in Theorem 1 does not depend on the fixed state a that we choose each time.

2. Variance calculation and equivalent matrix forms

The quantities $g_a(i,j)$, which appear in Theorem 1, involve both the state of the system at time n and the first return time to a fixed state a. We can further analyze them and get a representation in terms of the transition matrix in the case of finite state space. The following proposition serves this purpose. We use I_a for the $(s-1)$-dimensional identity matrix, since as we did for P to get P_a, the matrix I_a results from I excluding both its a-column and a-row.