Statistics and Probability Letters 118 (2016) 163–170

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

The Monte Carlo computation error of transition probabilities

Adam Nielsen

Konrad-Zuse-Zentrum für Informationstechnik Berlin, Takustraße 7, 14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 29 December 2015
Accepted 14 June 2016
Available online 7 July 2016

Keywords:
Reversible Markov chain
Monte Carlo methods
Computation error
Measurable state space
Markov operator

ABSTRACT

In many applications one is interested to compute transition probabilities of a Markov chain. This can be achieved by using Monte Carlo methods with local or global sampling points. In this article, we analyze the error by the difference in the L^2 norm between the true transition probabilities and the approximation achieved through a Monte Carlo method. We give a formula for the error for Markov chains with locally computed sampling points. Further, in the case of reversible Markov chains, we will deduce a formula for the error when sampling points are computed globally. We will see that in both cases the error itself can be approximated with Monte Carlo methods. As a consequence of the result, we will derive surprising properties of reversible Markov chains.

1. Introduction

In many applications, one is interested to approximate the term

$$\mathbb{P}[X_1 \in B \mid X_0 \in A]$$

from a Markov chain (X_n) with stationary measure μ. A solid method to do this is by using a Monte Carlo method. In this article, we will give the exact error between this term and the approximated term by the Monte Carlo method. It will turn out that when we approximate the term with N trajectories with starting points distributed locally in A, then the squared error is exactly given by

$$\frac{1}{N} \cdot \left(\mathbb{E}_{\mu_A}[\mathbb{P}_x[\tilde{X}_1 \in B]^2] - \mathbb{E}_{\mu_A}[\mathbb{P}_x[\tilde{X}_1 \in B]]^2 \right)$$

where $\mu_A(B) := \mu(A \cap B)$ is the stationary distribution restricted on A and $(\tilde{X}_n)$ is the reversed Markov chain. If the Markov chain (X_n) is reversible and if we approximate the term with N trajectories with global starting points, then the squared error is given by

$$\frac{1}{N} \cdot \left(\frac{\mathbb{P}[X_2 \in A,\, X_1 \in B \mid X_0 \in A]}{\mathbb{P}[X_0 \in A]} - \mathbb{P}[X_1 \in B \mid X_0 \in A]^2 \right).$$

We even give the exact error for a more generalized term which is of interest whenever one wants to compute a Markov State Model of a Markov operator based on an arbitrary function space which has application in computational drug design (Schütte and Sarich, 2015). Further, we derive from the result some surprising properties for reversible Markov chains. For example, we will show that reversible Markov chains rather return to a set then being in a set, i.e.

$$\mathbb{P}[X_2 \in A \mid X_0 \in A] \geq \mathbb{P}[X_0 \in A]$$

for any measurable set A.

E-mail address: nielsen@zib.de.

http://dx.doi.org/10.1016/j.spl.2016.06.011
0167-7152/

2. Basics

We denote with (E, Σ, μ) and $(\Omega, \mathcal{A}, \mathbb{P})$ probability spaces on any given sets E and Ω. We call a map $p\colon E \times \Sigma \to [0, 1]$ *Markov kernel* if

- $A \to p(x, A)$ is almost surely a probability measure on Σ and
- $x \to p(x, A)$ is measurable for all $A \in \Sigma$.

We denote with $(X_n)_n$, $X_n\colon \Omega \to E$ a Markov chain on a measurable state space (Revuz, 1984; Meyn and Tweedie, 2009), i.e. there exists a Markov kernel p with

$$\mathbb{P}[X_0 \in A_0,\ X_1 \in A_1, \ldots, X_n \in A_n] = \int_{A_0} \cdots \int_{A_{n-1}} p(y_{n-1}, A_n)\, p(y_{n-2}, dy_{n-1}) \ \ldots\ p(y_0, dy_1)\, \mu(dy_0)$$

for any $n \in \mathbb{N},\ A_0, \ldots, A_n \in \Sigma$. In this article, we only need this formula for $n \leq 2$, where it simplifies to

$$\mathbb{P}[X_0 \in A_0,\ X_1 \in A_1, X_2 \in A_2] = \int_{A_0} \int_{A_1} p(y, A_2)\, p(x, dy)\, \mu(dx)$$

for any $A_0, A_1, A_2 \in \Sigma$. We assume throughout this paper that μ is a stationary measure, i.e. $\mathbb{P}[X_n \in A] = \mu(A)$ for any $n \in \mathbb{N},\ A \in \Sigma$.

We call a Markov chain (X_n) *reversible* if

$$\int_A p(x, B)\, \mu(dx) = \int_B p(x, A)\, \mu(dx)$$

holds for any $A, B \in \Sigma$. This is equivalent to

$$\mathbb{P}[X_0 \in A,\ X_1 \in B] = \mathbb{P}[X_0 \in B,\ X_1 \in A]$$

for any $A, B \in \Sigma$. In particular, it implies that μ is a stationary measure. We denote with $L^1(\mu)$ the space of all μ-integrable functions (Bauer, 1992, Page 99). It is known (Hopf, 1954, Theorem 2.1) that there exists a Markov operator $P\colon L^1(\mu) \to L^1(\mu)$ with

$$\|Pf\|_{L^1} = \|f\|_{L^1} \quad \text{and} \quad Pf \geq 0$$

for all non-negative $f \in L^1(\mu)$ which is associated to the Markov chain, i.e.

$$\int_E p(x, A)\, f(x)\, \mu(dx) = \int_A (Pf)(x)\, \mu(dx)$$

for all $A \in \Sigma,\ f \in L^1(\mu)$. Since μ is a stationary measure, we have $P(L^2(\mu)) \subset L^2(\mu)$ and we can restrict P onto $L^2(\mu)$ (Baxter and Rosenthal, 1995, Lemma 1). It is known (Hopf, 1954) that a reversed Markov chain $(\tilde{X}_n)$ exists with

$$\mathbb{P}[X_1 \in B, X_0 \in A] = \mathbb{P}[\tilde{X}_1 \in A, \tilde{X}_0 \in B]$$

and that the Markov operator can be evaluated point-wise as

$$Pf(x) = \mathbb{E}_x[f(\tilde{X}_1)]. \tag{1}$$

In the case where (X_n) is reversible, we have that the Markov operator $P\colon L^2(\mu) \to L^2(\mu)$ is self-adjoint (Huisinga, 2001, Proposition 1.1) and that $X_n = \tilde{X}_n$, thus it can be evaluated point-wise by

$$Pf(x) = \mathbb{E}_x[f(X_1)] \tag{2}$$

Throughout the paper we note for any probability measure μ the expectation as

$$\mathbb{E}_\mu[f] = \int_E f(x)\, \mu(dx)$$

and use the shortcut $\mathbb{E} := \mathbb{E}_{\mathbb{P}}$. We denote with $\langle f, g\rangle_\mu = \int_E f(x)\, g(x)\, \mu(dx)$ the scalar product on $L^2(\mu)$. In this article, we are interested in the quantity

$$C := \frac{\langle f, Pg\rangle_\mu}{\langle f, 1\rangle_\mu} \tag{3}$$

where 1 denotes the constant function that is everywhere one and $f, g \in L^2(\mu)$ with $\langle f, 1\rangle_\mu \neq 0$. For the special case where $f(x) = 1_A(x)$ and $g(x) = 1_B(x)$ holds, we obtain

$$C = \frac{1}{\mu(A)} \int_B p(x, A)\, \mu(dx) = \mathbb{P}[X_1 \in B \mid X_0 \in A].$$

There are many applications that involve an approximation of C for different types of functions f, g which can be found in Schütte and Sarich (2015).