Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

ADAM NIELSEN

The Monte Carlo Computation Error of Transition Probabilities

ZIB Report 16-37 (July 2016)

Zuse Institute Berlin
Takustr. 7
D-14195 Berlin

Telefon: +49 30-84185-0
Telefax: +49 30-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

The Monte Carlo Computation Error of Transition Probabilities

Adam Nielsen

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 in the L^2 norm 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 in the L^2 norm 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 [10, 11]. 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.

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 [8, 6], i.e.there exists a Markov kernel p with

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

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 [1, Page 99]. It is known [3, 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 $L^2(\mu)$ [2, Lemma 1]. It is known [3] 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 [4, 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 [10].

3 Computation of the Local Error

To compute the error locally, we rewrite

$$C = \frac{\langle f, Pg\rangle_\mu}{\langle f, 1\rangle_\mu} = \int Pg(x)\,\mu_i(dx)$$

with

$$\mu_i(A) = \frac{1}{\langle f, 1\rangle_\mu} \int_A f(x)\mu(dx).$$

This term can be approximated by Monte Carlo methods [5]. To do so, we need random variables $Y, Y_1, \ldots, Y_N\colon \Omega \to E$ distributed according to μ_i. Then we can approximate the term C by

$$\tilde{C} := \frac{1}{N} \sum_{k=1}^{N} Pg(Y_k).$$

The error can be stated as

$$||C - \tilde{C}||^2_{L^2(\mathbb{P})} = \frac{\mathrm{VAR}[Pg(Y)]}{N}.$$

This follows from $\mathbb{E}[\tilde{C}] = C$ and

$$\begin{aligned} ||C - \tilde{C}||^2_{L^2(\mathbb{P})} &= \mathrm{VAR}[\tilde{C}] \\ &= \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{VAR}[Pg(Y)] \\ &= \frac{\mathrm{VAR}[Pg(Y)]}{N}. \end{aligned}$$

In this case the variance is simply given as

$$\begin{aligned} \mathrm{VAR}[Pg(Y)] &= \mathbb{E}[Pg(Y)^2] - \mathbb{E}[Pg(Y)]^2 \\ &= \mathbb{E}_{\mu_i}[\mathbb{E}_x[g(\tilde{X}_1)]^2] - \mathbb{E}_{\mu_i}[\mathbb{E}_x[g(\tilde{X}_1)]]^2. \end{aligned}$$

In the case where the Markov chain (X_n) is reversible, the variance simplifies to

$$\mathrm{VAR}[Pg(Y)] = \mathbb{E}_{\mu_i}[\mathbb{E}_x[g(X_1)]^2] - \mathbb{E}_{\mu_i}[\mathbb{E}_x[g(X_1)]]^2.$$

If further $g = 1_A$, we obtain

$$\mathrm{VAR}[Pg(Y)] = \mathbb{E}_{\mu_i}[\mathbb{P}_x[X_1 \in A]^2] - \mathbb{E}_{\mu_i}[\mathbb{P}_x[X_1 \in A]]^2.$$

3.1 Example

Consider the solution $(Y_t)_{t \geq 0}$ of the stochastic differential equation

$$dY_t = -\nabla V(Y_t) + \sigma dB_t$$

with $\sigma > 0$ and $V(x) = (x-2)^2(x+2)^2$. The potential V and a realization of (Y_t) are shown in Figure 1. The solution (Y_t) is known to be reversible, see [7, Proposition 4.5]. For any value $\tau > 0$ the family $(X_n)_{n \in \mathbb{N}}$ with $X_n := Y_{n \cdot \tau}$ is a reversible Markov chain. Let us partition $[-3, 3]$ into twenty equidistant sets $(A_l)_{l=1,\ldots,20}$. We compute a matrix $M \in \mathbb{R}^{20 \times 20}$ with

$$M(i,j) = \mathbb{E}_{\mu_i}[\mathbb{P}_x[X_1 \in A_j]^2] - \mathbb{E}_{\mu_i}[\mathbb{P}_x[X_1 \in A_j]]^2$$

with $\mu_i(B) = \mu(A_i \cap B)$. In each set A_i we sample points $x_1^i, \ldots, x_N^i$ with the Metropolis Monte Carlo Method distributed according to μ_i. For each point x_j^i we sample M trajectories starting in x_j^i with endpoint $y_{k,j}^i$ for $k = 1, \ldots, M$. We then approximate

$$\mathbb{E}_{\mu_i}[\mathbb{P}_x[X_1 \in A_j]^2] \approx \frac{1}{N} \sum_{l=1}^{N} \left(\frac{1}{M} \sum_{k=1}^{M} 1_{A_j}(y_{k,l}^i) \right)^2$$

and

$$\mathbb{E}_{\mu_i}[\mathbb{P}_x[X_1 \in A_j]]^2 \approx \left(\frac{1}{N} \sum_{l=1}^{N} \frac{1}{M} \sum_{k=1}^{M} 1_{A_j}(y_{k,l}^i) \right)^2$$

The local sampled points and the matrix M are shown in Figure 2. It shows that in order to compute the transition probability from 0 to 2 or from 0 to -2, many trajectories are needed.

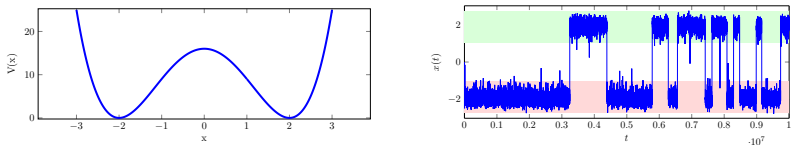

Figure 1: Potential (left) and trajectory (right).

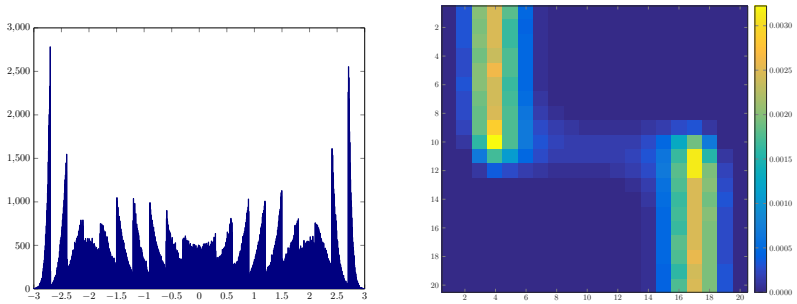

Figure 2: Local starting points (left) and matrix M (right).

4 Computation of the Global Error

We assume in this section that (X_n) is reversible.

If we denote with

$$\phi(x) := Pf(x) \cdot g(x) \cdot \frac{1}{\langle f, 1\rangle_\mu},$$

then one can rewrite

$$C = \frac{\langle f, Pg\rangle_\mu}{\langle f, 1\rangle_\mu} = \frac{\langle Pf, g\rangle_\mu}{\langle f, 1\rangle_\mu} = \int_E \phi(x)\, \mu(dx).$$

Analogous to the local case, we can approximate C by Monte Carlo methods. However, this time we need random variables $Y, Y_1, \ldots, Y_N : \Omega \to E$ distributed according to μ. Then we can approximate the term C by

$$\tilde{C} := \frac{1}{N} \sum_{k=1}^{N} \phi(Y_k).$$

Again, the error can be stated as

$$||C - \tilde{C}||^2_{L^2(\mathbb{P})} = \frac{\mathrm{VAR}[\phi(Y)]}{N}.$$

The main result of the article is the following theorem:

Theorem 1 *It holds*

$$\mathrm{VAR}[\phi(Y)] = \frac{\mathbb{E}_{\nu_f}[g^2(X_1)\, f(X_2)]}{\mathbb{E}[f(X_0)]} - \mathbb{E}_{\nu_f}[g(X_1)]^2$$

with

$$\nu_f(A) = \int_A \frac{f(X_0(\omega))}{\int_\Omega f(X_0(\tilde{\omega}))\, \mathbb{P}(d\tilde{\omega})}\, \mathbb{P}(d\omega).$$

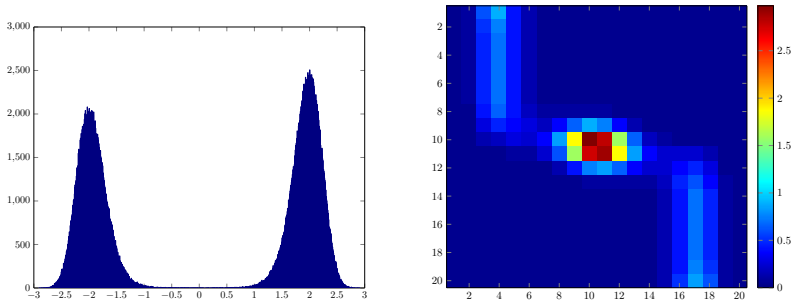

Figure 3: Global starting points (left) and matrix M (right).

The theorem shows that one can compute the exact error. Before we step into the proof, we will first show some applications of the theorem.

4.1 Example

In the case where $f(x) = 1_A(x)$, $g(x) = 1_B(x)$ the variance simplifies to

$$\mathrm{VAR}[\phi(Y)] = \frac{\mathbb{P}[X_2 \in A,\, X_1 \in B \mid X_0 \in A]}{\mathbb{P}[X_0 \in A]} - \mathbb{P}[X_1 \in B \mid X_0 \in A]^2.$$

We consider again the example from Section 3.1. This time, we compute starting points globally with the Metropolis Monte Carlo Method distributed according to μ and we compute the matrix

$$M_{ij} = \frac{\mathbb{P}[X_2 \in A_i,\, X_1 \in A_j \mid X_0 \in A_i]}{\mathbb{P}[X_0 \in A_i]} - \mathbb{P}[X_1 \in A_j \mid X_0 \in A_i]^2.$$

The results are shown in Figure 3. One may note that the entries of the global variance matrix are significantly higher then of the local variance matrix. The reason for that is the following. In the global case, the matrix $M(i,j)$ gives indication how many global trajectories one needs to keep the error of $\mathbb{P}[X_1 \in A_j \mid X_0 \in A_i]$ small. However, one can only use those trajectories with starting points in A_i. But in the local case, we only sample starting points in A_i and thus we do not need to dismiss any trajectories, this implies that we need less trajectories and thus it becomes clear that $M(i,j)$ is smaller in the local case. Especially, the global scheme has difficulties to capture the transition probability from a transition region (here the area near 0) to a transition region. Which is in turn no challenge for the local scheme.

5 Inequalities for Reversible Markov Chains

Since $\mathrm{VAR}[\phi(Y)] \geq 0$, we obtain

$$\frac{\mathbb{E}_{\nu_f}[g^2(X_1)\, f(X_2)]}{\mathbb{E}[f(X_0)]} \geq \mathbb{E}_{\nu_f}[g(X_1)]^2$$

for any $f, g \in L^2(\mu)$. If we set $g(x) = 1$, we obtain

$$\mathbb{E}_{\nu_f}[f(X_2)] \geq \mathbb{E}[f(X_0)].$$

For the case where $f(x) = 1_A$ for $A \in \Sigma$, this simplifies to

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

In short: Reversible Markov chains prefer to return to a set instead of being there. If $E = \{1, \ldots, n\}$ is finite and the Markov chain is described by a transition matrix $P \in \mathbb{R}^{n \times n}$ which is reversible according to the probability vector $\pi \in \mathbb{R}^n$, i.e.

$$\pi_i p_{ij} = \pi_j p_{ji},$$

then the inequality states

$$P^2(i,i) \geq \pi_i.$$

This shows by the way that from

$$\sum_{i=1}^{n} P^2(i,i) = 1$$

it immediately follows

$$P^2(i,i) = \pi_i.$$

Another inequality can be obtained as follows. Fix some sets $A, B \in \Sigma$ with $A \cap B = \emptyset$. Define the function f at point $x \in E$ to be the probability that the Markov chain (X_n) hits next set A before hitting set B when starting in x. Since the Markov chain is reversible, this is equivalent to the probability that one visited last set A instead of set B. This function is also known as committor function [9]. The term $\mathbb{E}_{\nu_f}[f(X_2)]$ is then given as the conditioned probability that one came last from A instead of B, moves for two steps, and returns then next to A instead of B. The probability for this event is thus always higher, then $\mathbb{E}[f(X_0)]$ which represents simply the probability that one came last from A instead of B.

Example

Consider the finite reversible Markov chain

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

visualized in Figure 4 with stationary measure $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. In this case we have

$$P^2(i,i) = \frac{1}{2} \geq \frac{1}{3} = \pi_i.$$

The probability that one came last from $A := \{1\}$ instead of $B := \{2\}$ is given by

$$\mathbb{E}[f(X_0)] = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}.$$

The conditioned probability that one came last from A instead of B, moves for two steps, and returns then next to A instead of B is given by

$$\mathbb{E}_{\nu_f}[f(X_2)] = \frac{2}{3} \cdot \left(\frac{1}{4} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} + \frac{1}{4} \cdot 0\right) + 0 + \frac{1}{3} \cdot \left(\frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot 0\right) = \frac{7}{12}$$

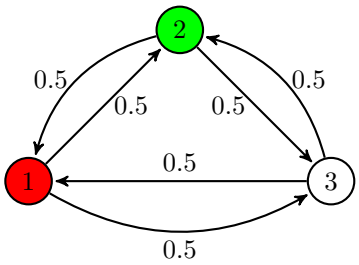

Figure 4: Visualized simple Markov chain

where we have used that $\nu_f(\{1\}) = \frac{2}{3}$, $\nu_f(\{2\}) = 0$ and $\nu_f(\{3\}) = \frac{1}{3}$. Thus, we have as predicted

$$\frac{7}{12} = \mathbb{E}_{\nu_f}[f(X_2)] \geq \mathbb{E}[f(X_0)] = \frac{1}{2}.$$

6 The proof

We will now give the proof of Theorem 1. First, note that we have

$$\mathrm{VAR}[\phi(Y)] = \mathbb{E}[\phi^2(Y)] - \mathbb{E}[\phi(Y)]^2.$$

For sets $A, B \in \Sigma$, we have

$$\mathbb{E}[\mathbb{1}_A(X_0)\,\mathbb{1}_B(X_1)] = \mathbb{P}[X_0 \in A, X_1 \in B] = \int_E \mathbb{1}_A(x) \int_E \mathbb{1}_B(y)\, p(x, dy)\, \mu(dx).$$

For measurable functions $f, g \geq 0$, we then obtain

$$\mathbb{E}[f(X_0)\, g(X_1)] = \int_E f(x) \int_E g(y)\, p(x, dy)\, \mu(dx). \tag{4}$$

For $\phi(x) = Pf(x)\, g(x) \frac{1}{\langle f, \mathbb{1}\rangle_\mu}$ we first compute

$$\begin{aligned}
\mathbb{E}[\phi^2(Y)] &= \frac{1}{\langle f, \mathbb{1}\rangle_\mu^2} \int_E \big(Pf(x)\, g(x)\big)^2\, \mu(dx) \\
&= \frac{1}{\langle f, \mathbb{1}\rangle_\mu^2} \int_E Pf(x) \Big(g^2 \cdot Pf\Big)(x)\, \mu(dx) \\
&= \frac{1}{\langle f, \mathbb{1}\rangle_\mu^2} \int_E f(x) \Big(P\Big(g^2 \cdot Pf\Big)(x)\Big)\, \mu(dx) \\
&= \frac{1}{\langle f, \mathbb{1}\rangle_\mu^2} \int_E f(x) \left(\int_E g^2(y)\, Pf(y)\, p(x, dy)\right)\, \mu(dx)
\end{aligned}$$

leading to

$$\mathbb{E}[\phi^2(Y)] = \frac{1}{\langle f, \mathbb{1}\rangle_\mu^2} \int_E f(x) \int_E g^2(y)\, \mathbb{E}_y[f(X_1)]\, p(x, dy)\, \mu(dx),$$

where we have used Equation (2). From Equation (4) we get

$$\mathbb{E}[\phi^2(Y)] = \frac{1}{\langle f, \mathbb{1}\rangle_\mu^2} \mathbb{E}\Big[f(X_0)\, g^2(X_1)\, \mathbb{E}_{X_1}[f(X_1)]\Big].$$

Further, according to [6, Equation (3.28) in Chapter 3], we have

$$\mathbb{E}_{X_1}[f(X_1)] = \mathbb{E}[f(X_2) \mid \mathcal{F}_1],$$

where $\mathcal{F}_1 = \sigma(X_0, X_1)$. Thus, we obtain

$$\mathbb{E}\big[f(X_0)\, g^2(X_1)\, \mathbb{E}_{X_1}[f(X_1)]\big] = \mathbb{E}\big[f(X_0)\, g^2(X_1)\, \mathbb{E}[f(X_2) \mid \mathcal{F}_1]\big]$$

Also note that we have

$$\mathbb{E}[\mathbb{1}_A\, \mathbb{E}[f(X_2) \mid \mathcal{F}_1]] = \mathbb{E}[\mathbb{1}_A\, f(X_2)]$$

for any $A \in \mathcal{F}_1$. Since $f(X_0)\, g^2(X_1)$ is measurable according to $\mathcal{F}_1$, we obtain

$$\mathbb{E}\big[f(X_0)\, g^2(X_1)\, \mathbb{E}[f(X_2) \mid \mathcal{F}_1]\big] = \mathbb{E}\big[f(X_0)\, g^2(X_1)\, f(X_2)\big].$$

Thus we finally get

$$\mathbb{E}[\phi^2(Y)] = \mathbb{E}_{\nu_f}[g^2(X_1)\, f(X_2)]\, \mathbb{E}_\mu[f]^{-1}$$

with

$$\nu_f(A) = \int_A \frac{f(X_0(\omega))}{\int_\Omega f(X_0(\tilde{\omega}))\, \mathbb{P}(d\tilde{\omega})}\, \mathbb{P}(d\omega)$$

for all measurable sets A.

Finally, we have

$$\begin{aligned}
\mathbb{E}[\phi(Y)] &= \frac{1}{\mathbb{E}_\mu[f]} \int_E Pf(x)\, g(x)\, \mu(dx) \\
&= \frac{1}{\mathbb{E}_\mu[f]} \int_E f(x)\, Pg(x)\, \mu(dx) \\
&= \frac{1}{\mathbb{E}_\mu[f]} \int_E f(x) \left[\int_E g(y)\, p(x, dy)\right] \mu(dx) \\
&= \frac{1}{\mathbb{E}_\mu[f]} \mathbb{E}[f(X_0) g(X_1)] \\
&= \mathbb{E}_{\nu_f}[g(X_1)].
\end{aligned}$$

This gives us the exact formula for the variance.

Acknowledgement

The author gatefully thanks the Berlin Mathematical School for financial support.

References

[1] Heinz Bauer. *Maß und Integrationstheorie*. Walter de Gruyter, Berlin; New York, 2nd edition, 1992.

[2] John R. Baxter and Jeffrey S. Rosenthal. Rates of convergence for everywhere-positive markov chains. *Statistics & Probability Letters*, 22:333–338, 1995.

[3] Eberhard Hopf. *The general temporally discrete Markoff process*, volume 3. 1954.

[4] Wilhelm Huisinga. *Metastability of Markovian systems*. PhD thesis, Freie Universitt Berlin, 2001.

[5] David J. C. MacKay. Introduction to Monte Carlo methods. In M. I. Jordan, editor, *Learning in Graphical Models*, NATO Science Series, pages 175–204. Kluwer Academic Press, 1998.

[6] Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, second edition edition, 2009.

[7] Grigorios A. Pavliotis. *Stochastic Processes and Applications*. Springer, 2014.

[8] Daniel Revuz. *Markov Chains*. North-Holland Mathematical Library, 2nd edition, 1984.

[9] Christof Schütte and Marco Sarich. *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*. 2014. A co-publication of the AMS and the Courant Institute of Mathematical Sciences at New York University.

[10] Christof Schütte and Marco Sarich. A critical appraisal of markov state models. *The European Physical Journal Special Topics*, pages 1–18, 2015.

[11] Marcus Weber. *Meshless Methods in Conformation Dynamics*. PhD thesis, Freie Universitt Berlin, 2006.