

Monte Carlo simulation for estimating rare event probabilities and parameters in Markov process models

Doctoral Thesis

Author(s):

Amrein, Michael

Publication date:

2011

Permanent link:

<https://doi.org/10.3929/ethz-a-006383172>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Diss. ETH No. 19452

Monte Carlo simulation for estimating rare event probabilities and parameters in Markov process models

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
MICHAEL AMREIN
Dipl. Math. ETH
born November 27, 1980
citizen of Neudorf LU

accepted on the recommendation of
Prof. Dr. Hans Rudolf Künsch, examiner
Prof. Dr. Leonhard Held, co-examiner

2011

Abstract

Markov processes models appear in many different fields, for example in queueing theory, epidemic models, stochastic kinetics of chemical reactions or financial engineering. Often, simple algorithms exist to generate sample paths of this processes. This makes analysis based on simulation, so called Monte Carlo methods, very attractive. We study two specific estimation problems in this context.

First we consider a simulation technique, called importance splitting, to estimate very small entrance probabilities for Markov processes by splitting sample paths at various level surfaces of the so called importance function before reaching the set of interest. This can be done in many ways, yielding different variants of the method.

In this context, we propose a new one, called fixed number of successes. We prove unbiasedness for the new and some known variants, because in many papers, the proof is based on an incorrect argument. Further, we analyze its behavior in a simplified setting, which is appropriate if the importance function is well chosen, in terms of efficiency and asymptotics in comparison to the standard variant. The main difference is that the new variant controls the precision of the estimator rather than the computational effort. Our analysis and simulation examples show that it is rather robust in terms of parameter choice and we present a two-stage procedure which also yields confidence intervals.

The choice of the importance function which governs the placement

of the splitting surfaces is a major issue in splitting estimation. Thus we present and motivate importance functions when estimating the entrance probability of a continuous-time Markov chain on a discrete state space into some set, for example a buffer overflow in a tandem Jackson network, before hitting another set or during some finite time interval. We show that they yield good results in practice, especially combined with our two-stage procedure based on fixed number of successes.

Second, we present a simulation methodology for Bayesian estimation of rate parameters in Markov jump processes arising for example in stochastic kinetic models. To handle the problem of missing components and measurement errors in observed data, we embed the Markov jump process into the framework of a general state space model. Markov chain Monte Carlo and particle filter type algorithms are introduced, which allow sampling from the posterior distribution of the rate parameters and the Markov jump process, also in data-poor scenarios. The algorithms are illustrated by applying them to rate estimation in a model for prokaryotic auto-regulation and in the stochastic Oregonator, respectively.

Zusammenfassung

Markov-Prozesse werden in den unterschiedlichsten Bereichen verwendet, zum Beispiel in der Warteschlangentheorie, in epidemischen Modellen, in der stochastischen Reaktionskinetik oder im “Financial Engineering”. Meist existieren einfache Algorithmen, um die entsprechenden Zufallstrajektorien zu generieren. Dies macht Analysen basierend auf Simulationen, sogenannte Monte-Carlo-Methoden, besonders attraktiv. Wir studieren in dieser Dissertation zwei Probleme in diesem Zusammenhang.

Zuerst betrachten wir eine Simulationstechnik zur Schätzung von kleinen Eintrittswahrscheinlichkeiten einer Markov-Kette in eine bestimmte Menge ihres Zustandsraumes namens “importance splitting”. Dabei wird die Simulation der Trajektorien nach Erreichen einer gewissen Stufe, definiert über die sogenannte “importance function”, immer wieder neu gestartet, bis die nächste erreicht ist. Für dieses Aufsplitten existieren verschiedene Varianten.

In diesem Zusammenhang präsentieren wir eine neue Variante namens “fixed number of successes”. Wir beweisen Erwartungstreuung für diese, aber auch für einige schon bekannte Varianten, da der Beweis in einigen Publikationen ein falsches Argument enthält. Weiter analysieren wir sie unter einer vereinfachenden Annahme – welche angemessen ist, wenn die “importance function” gut gewählt ist – in Bezug auf Effizienz und Asymptotik im Vergleich zur Standard-Variante. Der Hauptunterschied ist, dass unsere neue Variante die Präzision (im Sin-

ne der Varianz relativ zur quadrierten Schätzgröße) kontrolliert und nicht den Berechnungsaufwand. Weiter zeigen unsere Analyse und unsere Simulationsbeispiele, dass sie robust in Bezug auf die Wahl ihrer Simulations-Parameter ist. Basierend auf unseren Ergebnissen haben wir eine zweistufige Schätzprozedur entwickelt, die auch Vertrauensintervalle liefert.

Die Wahl der “importance function” ist eines der Hauptprobleme in der Anwendung von “importance splitting”. Wir präsentieren solche Funktionen wenn die Eintrittswahrscheinlichkeit einer zeitstetigen Markov-Kette auf diskretem Zustandsraum in eine bestimmte Menge, entweder bevor Eintritt in eine andere Menge oder bevor Ablauf einer gewissen Zeit, geschätzt werden soll. Weiter zeigen wir, dass diese Funktionen gute Resultate in der Praxis liefern, speziell wenn sie mit unserer zweistufigen Schätzprozedur basierend auf der Variante “fixed numbers of successes” kombiniert werden.

Zweitens stellen wir in dieser Dissertation eine Simulationsmethodologie zur Bayesschen Schätzung von Raten-Parametern in Markov-Sprung-Prozessen vor, die zum Beispiel in der stochastischen Reaktionskinetik vorkommen. Um fehlende Werte oder Messfehler ins Modell zu integrieren, betten wir den Markov-Sprung-Prozess in ein “general state space model” ein. Markov-Ketten-Monte-Carlo- und Partikel-Filter-Methoden werden zur Simulation von der A-posteriori-Verteilung der Parameter und des Markov-Sprung-Prozesses benützt, auch in Situationen, in denen nur wenig Daten zur Verfügung stehen. Die Algorithmen werden anhand von Anwendungen zur Schätzung der Reaktionsraten in einem einfachen Modell für prokaryotische Auto-Regulation und im stochastischen Oregonator illustriert.