Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 119 (2009) 491–517
stochastic processes and their applications
www.elsevier.com/locate/spa

Importance sampling for a Markov modulated queuing network

Ali Devin Sezer*

Middle East Technical University, Institute of Applied Mathematics, Ankara, Turkey

Received 20 September 2007; received in revised form 19 February 2008; accepted 20 February 2008
Available online 6 March 2008

Abstract

Importance sampling (IS) is a variance reduction method for simulating rare events. A recent paper by Dupuis, Wang and Sezer [Paul Dupuis, Ali Devin Sezer, Hui Wang, Dynamic importance sampling for queueing networks, Annals of Applied Probability 17 (4) (2007) 1306–1346] exploits connections between IS and stochastic games and optimal control problems to show how to design and analyze simple and efficient IS algorithms for various overflow events of tandem Jackson Networks. The present paper carries out a program parallel to the paper by Dupuis et al. for a two node tandem network whose arrival and service rates are modulated by an exogenous finite state Markov process. The overflow event we study is the following: the number of customers in the system reaches n without the system ever becoming empty, given that initially the system is empty.

MSC: 60K25; 49N90 (60F10, 65C05, 91A15, 91A23)

Keywords: Dynamic importance sampling; Rare event simulation; Tandem queues; Queuing networks; Markov modulated; Regime switch; Overflow probability; Large deviations; Isaacs equation; Optimal control

1. Introduction

Importance sampling (IS) is a variance reduction method for simulating rare events. The idea in IS is to change the sampling distribution (and modify the Monte Carlo estimator accordingly) to reduce estimator variance.

* Tel.: +90 312 210 5608; fax: +90 312 210 2985.
E-mail address: devin@metu.edu.tr.

0304-4149/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2008.02.009

In the present paper we develop an asymptotically optimal IS algorithm[1] for two tandem queues whose service and arrival processes are modulated by a finite state Markov chain. The system is assumed to be stable in the sense that the average service rate at each node is greater than the average arrival rate to that node, see Assumption 2. One commonly used quantity to measure the performance of systems such as this one is the following probability: given that initially the network is empty, the total number of customers in the network becomes n before the network empties [12,21]. Under the stability assumption and for large n one would expect this event to be rare. In the present paper we will be interested in the IS estimation of this rare event.

The iid version of this problem (constant arrival and service rates) was first posed in [21] in 1989. The discontinuous dynamics of the queuing process near the boundaries of its state space (i.e., when few customers remain in some of the nodes) makes the IS construction problem difficult for any queuing network including the case of two tandem nodes [18,12]. This property rules out iid sampling distributions (such as those developed in [20] in the context of a random walk on the real line and in [21] in the context of two tandem Jackson nodes) as candidates for efficient IS samplers and forces one to search for a good IS distribution among dynamic (i.e., Markovian) distributions. For a more in depth discussion of these issues we refer the reader to [18,17,12,4].

An asymptotically optimal IS algorithm with optimality proofs for the iid case was first developed in [18] using the "game approach" to IS [4,5,7]. This approach was first introduced in [4] in the context of the estimation of Cramer's theorem type small probabilities. In the game approach, one derives an Isaacs equation and a set of boundary conditions from a dynamic game interpretation of the problem. Appropriate smooth subsolutions to these generate asymptotically optimal and easily implementable dynamic IS schemes. These schemes are dynamic (or Markovian) in the sense that the IS transition probabilities depend on the current state of the process being sampled (in the case of queuing networks the state of the process at any time is the number of customers at each node at that time). Recent work on the game approach to IS includes [5,7,6] and in the context of queuing networks [17,18].

In many applications, perhaps the most popular method of extending the dynamics of an iid increments model is to introduce Markov modulation. To the best of our knowledge, there is not even a heuristic IS algorithm available in the current literature for queuing networks with Markov modulated rates. Our first goal in this paper is to provide the first optimal IS algorithms for estimating overflow events of a queuing model involving this type of very common and important dynamics. Our second goal is to provide support for the idea that the game approach provides a systematic and versatile framework for the construction of IS schemes for a wide range of stochastic processes. As we discuss below, the extension of the game approach to the MM dynamics requires several new ideas. But the main idea remains the same, which we repeat: the IS problem for rare events of a Markov process which decay exponentially in a parameter can be represented as a sequence of dynamic stochastic games and a limit Isaacs equation and boundary conditions can be derived. Appropriate subsolutions to these can be used to construct optimal IS schemes.

The reason why we treat only the two tandem nodes case is the following. As we point out in the following paragraphs, there are some nontrivial features of an extension to the MM setup and we believe that these can be communicated in the context of two tandem nodes. In our opinion,

[1] Or an IS *scheme*. There seems to be no formal definition of this term in the literature. In the present paper we use the words 'scheme' and 'algorithm' interchangeably.