Fast simulation of Markov fluid models in conjunction with large deviations

Mandjes, M.

1993

aculteit der Economische Wetenschappen en Econometrie
ET akgroep Econometrie

SERIE RESEARCH MEMORANDA

Fast Simulation of Markov Fluid Models in conjunction with Large Deviations

Michel Mandjes

Research Memorandum 1993-58

december 1993

vrije Universiteit *amsterdam*

Fast Simulation of Markov Fluid Models in conjunction with Large Deviations

Michel Mandjes
Department of Econometrics
Vrije Universiteit Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands

November 23, 1993

Abstract

In this paper Markov fluid models with buffer overflows are studied. Our aim is to find variance reduction for estimating probabilities of rare events via Monte Carlo simulation. On different probabilistic levels, we deduce large deviations expressions for the distribution of the underlying Markov chain in case of the buffer contents reaching a high level. Then we shall consider variance reduction by means of importance sampling where the new statistical law of the process is derived from the optimizing arguments of some large deviations expressions.

• Finite-capacity Markov Fluid models • Level 1, 2, and 3 Large Deviations • Entropy functions • Monte Carlo Simulation • Importance Sampling

1 Introduction

IN THIS PAPER we examine the probabilistic behavior of finite capacity Markov fluid models. Instead of concentrating on statistics connected with the steady state of the buffer occupancy, we investigate merely the deviant behavior, particularly the rare event of a buffer overflow. Our analysis is based on large deviations techniques. Recently, much attention is paid to this approach of analyzing fluid systems, see e.g. Weiss [1986], Ridder and Walrand [1992], and De Veciana *et al.* [1993]. For the fundamentals of large deviations theory we refer to Ellis [1985], Bucklew [1990], and Dembo and Zeitouni [1993].

Consider a continuous time Markov chain $X(\cdot)$. $X(\cdot)$ is a mapping from $[0,\infty)$ to a finite state space E, consisting of a certain finite number of states, say N, enumerated $1,\ldots,N$. Let $\Lambda := (\lambda_{ij})_{i,j=1}^{N}$ be the transition matrix of the chain, i.e., λ_{ij} is the infinitesimal transition rate from state i to j. By convention

$$\lambda_{ii} := -\sum_{j\neq i} \lambda_{ij} =: -\lambda_i.$$

Suppose $X(t) = i$ for some $t \in [0,\infty)$. Then the process stays in this state for a period which is exponentially distributed with mean λ_i^{-1} and makes a transition to state j with probability λ_{ij}/λ_i. We assume that it is possible to reach any states j from a different state i (where $i,j \in E$). So, the Markov chain is irreducible, and (since its state space is finite) consisting of merely positive recurrent states. Under this assumption, it can be proved that the chain has a unique steady state distribution, which can be calculated easily using the 'rate in = rate out'-principle. We call this distribution π_Λ.

Now consider a fluid system in a buffer with finite but 'large' capacity B. If the Markov chain is in state i, the buffer is filled continuously by an input stream at a constant rate of $\tilde{r}(i) > 0$ packets or data units per unit time. The buffer is emptied by a continuous output flow with rate $c > 0$, uniformly for all states. We denote by $r(i) := \tilde{r}(i) - c$ the net input per unit time in state i. If the buffer is empty and the process is in a state i with negative rate $r(i)$, the buffer remains empty; in case of a full buffer and $X(\cdot)$ being in a state i with $r(i) > 0$, the buffer contents remains B. To guarantee the stability of this system, we assume throughout $\langle \pi_\Lambda, r\rangle < 0$, where $\langle \cdot, \cdot \rangle$ denotes the inner product. To avoid trivialities, we suppose that there exists at least one state with a positive rate $r(\cdot)$. We call the resulting model a *finite capacity Markov modulated fluid model.*

Defining $S(t)$ as the amount of fluid in the buffer at time t (where $t \geq 0$), the development of the buffer contents is given by the following differential equation:

$$S(t + \mathrm{d}t) = S(t) + r(X(t))\,\mathrm{d}t,$$

whenever the right-hand side is feasible: $0 \leq S(t) + r(X(t))\,\mathrm{d}t \leq B$. Otherwise the buffer contents remains on equal level (i.e., 0 or B): $S(t + \mathrm{d}t) = S(t)$.

In a way described in Ridder and Walrand [1992] and Ridder [1993] we may recognize cycles in the evolution of $S(\cdot)$, consisting of idle as well as busy parts. In the remainder, we try to give rough estimates for the probability of an arbitrary cycle being an overflow cycle. In case of capacity B, we denote this probability by $\alpha(B)$ throughout. Notice that the cycles are not strictly regenerative, since the busy period may start in any state i with positive rate $r(i)$. How to overcome this difficulty is explained in the second example of section 8. Furthermore, it is clear that the underlying Markov chain does not behave according to the transition rates of Λ during an overflow: $S(\cdot)$ hits $[B, \infty)$ although $\langle \pi_\Lambda, r \rangle < 0$. In case of $B \to \infty$ we show that this underlying chain acts like another Markov chain, if in a cycle level B is reached (for large B) and we present a method to find the transition rates of this Markov chain.

As we saw before, the Markov modulated fluid model can be considered as a queueing system with finite buffer B that is filled up (continuously) by input streams and emptied at a rate c. Fluid models are proposed for modeling buffer behavior in switches of high speed communication networks. Important references in this field are Anick *et al.* [1982], Courcoubetis *et al.* [1991], Mitra [1988] and Stern and Elwalid [1991].

The main idea is that traffic comes mainly in bursts (or peaks) alternating with more quiet moments. In other words: we assume that the input of the queueing system is highly correlated in time. This means that the fluid model is a more natural representation of reality than models with 'homogeneous' input, e.g. Poisson processes. During a burst the packets arrive (almost) constantly in time rather than due to some random time process. Each potential customer of the network is associated with a suitable Markov chain for the time durations of the alternating traffic states (peak, quiet, ...) and with a suitable rates $\tilde{r}(i)$ for the generation of packets that the customer wants to send through the network (in each state i of the chain). When several customers or connections share the same buffer, we may superimpose their chains and rate functions into one chain and rate function, see example 3 of section 8.

In this paper the problem of finding rough asymptotic expressions for the probability $\alpha(B)$ generates a lot of useful 'by-products'. We will show that $\alpha(B)$ emerges as the result of the optimization of some large deviations expressions. We show that the optimizing arguments of these expressions provide us information about the probabilistic behavior of the buffer contents if conditioned on an overflow. In this way, we may find the transition rates of the chain if the fluid model reaches overflow. These 'new' rates are extremely useful, because they enable us to obtain variance reduction in Monte Carlo simulation, as explained later on.

The organization of this paper is as follows. In section 2 attention is paid to some topics related to the level 1 large deviations asymptotics of the loss probability. Section 3

deals with the level 2 approach. It is explained why this method is in general not useful for variance reduction objectives. The disadvantages of the level 2 approach appear not to apply to the level 3 approach and for that reason the latter is to be preferred. In section 4 the derivation of the level 3 large deviation rate function is sketched. The decay rate of $\alpha(B)$ appears to be the solution to an optimization problem. In section 5 this optimization is studied and necessary conditions for the optimizing arguments are given. Linking the level 1 and level 3 approach, it can be proven that these conditions are also sufficient, as can be found in section 6. The remainder of the paper treats the implementation of Importance Sampling in the fluid model in conjunction with large deviations, which facilitates accurate estimates of $\alpha(B)$. In the final section some simulation results are given, e.g. in the multiple source case.

2 Some notes on level 1 Large Deviations

TACKLING THE PROBLEM of finding asymptotics for $\alpha(B)$ we may investigate the free buffer process, just like in Ridder and Walrand [1992], i.e., $\bar{S}(t)$ denotes the amount of fluid in the system after removing the boundaries 0 and B. They fix an initial state i and let $\xi_j^{(i)}$ be the increment of fluid between the $(j-1)$th and jth return to state i; write $\xi^{(i)}$ for the generic variable. Defining the large deviation rate function or Legendre Fenchel transform by

$$I_1^{(i)}(x) := \sup_\theta \left(\theta x - \log \mathsf{E}\exp(\theta \xi^{(i)})\right),$$

it appeared that the decay rate of the loss probability satisfies

$$\lim_{B\to\infty} \frac{1}{B} \log \alpha(B) = \inf_{T>0} T\, I_1^{(i)}\left(\frac{1}{T}\right) = -\theta_i, \tag{1}$$

where θ_i denotes a positive solution of $\mathsf{E}\exp(\theta\xi^{(i)}) = 1$. The renewal reward theorem says that $\langle \pi_\Lambda, r\rangle$ equals the ratio of $\mathsf{E}\xi^{(i)}$ and $\mathsf{E}Y^{(i)}$, where $Y^{(i)}$ denotes the time elapsed between two consecutive returns to state i. We see that $\xi^{(i)}$ obviously has a negative mean. Now it can be proved that under very weak conditions, this equation has a unique positive root, see Mandjes [1993].

• Let the sequence $\{\xi_j^{(i)}, j \in \mathbb{N}\}$ consist of the increments of the free buffer process between two consecutive returns to state i, where $i \in E$. We let θ_i denote the positive solution for θ of $\mathsf{E}\exp(\theta\xi^{(i)}) = 1$. It is intuitively reasonable that $\theta_i \equiv \theta^*$ for all $j \in E$. Using large deviations it is easy to prove this property. This can be done as follows.

First define the following partial sums: $S_j^{(i)} := \sum_{k=1}^{j} \xi_k^{(i)}$, where $j \in \mathbb{N}$ and $i \in E$. Let $\beta^{(i)}(B)$ be the probability of the random walk $\{S_j^{(i)}, j \in \mathbb{N}\}$ reaching $[B,\infty)$, whereas $\alpha^{(i)}(B)$ is the probability that $[B,\infty)$ is hit before $(-\infty, 0]$ by $\{S_j^{(i)}, j \in \mathbb{N}\}$. *Slow Markov*

walk theory yields that under weak conditions it holds that for any i in E

$$\lim_{B\to\infty}\frac{1}{B}\log\beta^{(i)}(B)=\inf_{f}\int_0^{T(f)}I_1^{(i)}(f'(t))\mathrm{d}t, \tag{2}$$

where f ranges over all functions from $[0,\infty)$ to $[0,\infty)$, continuous and differentiable and hitting 1 (at time $T(f)>0$). As can be found in Bucklew [1990], the minimizing f is a straight line from the origin to level 1. Also,

$$\lim_{B\to\infty}\frac{1}{B}\log\alpha^{(i)}(B)=\inf_{g}\int_0^{T(g)}I_1^{(i)}(g'(t))\mathrm{d}t, \tag{3}$$

g ranging over all functions from $[0,\infty)$ to $[0,\infty)$, continuous and differentiable and hitting 1 before 0 (at $t=T(g)$). Since the optimizing f satisfies also the conditions for functions g we find that the large deviations expressions (2) and (3) are identical. It can be found in Bucklew [1990] that expression (2) equals $-\theta_i$.

Fix two distinct states in E, say i_1 and i_2. Define $F(i_1,i_2)$ as the smallest t such that $X(t)=i_2$, given that $X(0)=i_1$. Clearly, the following inequality holds:

$$\beta^{(i_1)}(B)\ \geq\ \mathsf{P}\big(\bar{S}(F(i_1,i_2))\geq 0\big)\times\beta^{(i_2)}(B)\times\mathsf{P}\big(\bar{S}(F(i_2,i_1))\geq 0\big).$$

Note that the first and last probability in the right-hand side of the previous display are positive constants independent of B. We call their product $C(i_1,i_2)$. The stated follows immediately:

$$-\theta_{i_1}=\lim_{B\to\infty}\frac{1}{B}\log\beta^{(i_1)}(B)\geq\lim_{B\to\infty}\frac{1}{B}\log C(i_1,i_2)+\lim_{B\to\infty}\frac{1}{B}\log\beta^{(i_2)}(B)=-\theta_{i_2}.$$

In a similar fashion: $-\theta_{i_1}\leq-\theta_{i_2}$. In view of the arbitrariness of i_1 and i_2 we have that $\theta_i\equiv\theta^*$, where $i\in E$. We found that the initial state i can be chosen arbitrarily in order to calculate large deviations expression (1).

• We saw that, without loss of generality, we have to find the unique positive root of $\mathsf{E}\exp(\theta\xi^{(1)})=1$. Conditioning on the jump from state 1, one may easily verify that the moment generating function can be written as follows

$$\begin{aligned}\mathsf{E}\exp(\theta\xi^{(1)}) &= \left(\frac{\lambda_{12}}{\lambda_1}\times\frac{\lambda_1}{\lambda_1-r(1)\theta}\times x_2\right)+\cdots+\left(\frac{\lambda_{1N}}{\lambda_1}\times\frac{\lambda_1}{\lambda_1-r(1)\theta}\times x_N\right)\\ &= \frac{\lambda_{12}x_2+\cdots+\lambda_{1N}x_N}{\lambda_1-r(1)\theta},\end{aligned}$$

where the variables x_i (where $i=2,\ldots,N$) denote the moment generating functions of the net amount of fluid generated by the free buffer model starting in state i until absorption in state 1. They are implicitly defined by

$$\begin{aligned}x_i &= \left(\frac{\lambda_{i1}}{\lambda_i-r(i)\theta}\right)+\sum_{j=2,j\neq i}^{N}\left(\frac{\lambda_{ij}}{\lambda_i}\times\frac{\lambda_i}{\lambda_i-r(i)\theta}\times x_j\right)\\ &= \frac{\lambda_{i1}+\lambda_{i2}x_2+\cdots+\lambda_{i,i-1}x_{i-1}+\lambda_{i,i+1}x_{i+1}+\cdots+\lambda_{iN}x_N}{\lambda_i-r(i)\theta}.\end{aligned}$$

So for a fixed θ, the moment generating function in this value can be found by solving a system of $N-1$ linear equations in $N-1$ unknowns, which can be done easily using standard Gauss-Jordan procedures. Applying for instance a bisection method we can find the value of θ^*.

Being able to compute θ^* as a function of Λ and $r = \tilde{r} - c$, we may find the value of the output rate such that the decay rate of $\alpha(B)$ equals a given prescribed value, again using a simple bisection technique. In this way, we computed the *effective bandwidth,* see De Veciana *et al.* [1993].

• In Mandjes [1993] is heuristically argued that under the assumption of a loss cycle the increments $\xi_j^{(1)}$ have an *exponentially twisted distribution.* This means the following: If the increments of the free buffer process between two visits to state 1 have distribution function $\mathsf{F}(\cdot)$, the increments have the following cumulative distribution function if conditioned on an overflow:

$$\int_{-\infty}^{u} e^{\theta x} \mathrm{d}\mathsf{F}(x) \quad \text{instead of} \quad \int_{-\infty}^{u} \mathrm{d}\mathsf{F}(x) = \mathsf{F}(u).$$

We call the generic variable of these increments $\xi^{(1)*}$. It is easy to check that the following 'translation property' holds for the moment generating function of the 'old' and 'new' increments:

$$\mathsf{E} \exp\left((\theta + \theta^*)\xi^{(1)}\right) = \mathsf{E} \exp\left(\theta \xi^{(1)*}\right), \tag{4}$$

So we found the moment generating function of the increments in case of an overflow. Under weak conditions the moment generating function is unique, and then we derived also its distribution. However, in general it is very difficult to find from this moment generating function transition rates $M = (\mu_{ij})_{i,j=1}^{N}$ of an underlying chain that go with $\xi^{(1)*}$. To gain some insight in the statistics of the Markov chain (under condition of the buffer contents hitting B) we may use the level 2 approach worked out by Ridder and Walrand [1992], which we shall review and discuss now.

3 The drawbacks of the level 2 approach

AS TREATED IN Ridder and Walrand [1992], we may find the in some sense most likely equilibrium distribution of the Markov chain under condition of the buffer contents reaching a high level.

In this level 2 approach it is considered what fraction of time the process stays in state i, where $i \in E$. Under the condition that the free buffer process reaches a high level, it is clear that there is a 'large discrepancy' between the resulting empirical distribution and the stationary distribution π_Λ. It is obvious that the inner product of this empirical distribution and r is positive, whereas $\langle \pi_\Lambda, r \rangle < 0$.

Instead of examining the roots of $\mathsf{E}\exp(\theta\xi^{(i)})$, we can find θ^* as a result of the optimization program

$$\inf T\, I_2(\pi),$$

where T ranges over $(0,\infty)$ and π over all probability measures on E such that the inner product of π and r is $1/T$. Here, the level 2 large deviation rate function is given by (Donsker and Varadhan [1975])

$$I_2(\pi) := -\inf_{u\gg 0}\sum_{i=1}^{N}\pi(i)\sum_{j=1}^{N}\frac{\lambda_{ij}u_j}{u_i},$$

where $u \gg 0$ means that the vector u is positive coordinatewise. It can be shown that the resulting θ^* equals expression (1). In Ridder [1993] is argued that, assuming that there exists an underlying Markov chain that is connected with $\xi^{(i)*}$, the equilibrium distribution of this chain equals the optimizing π, say π_M. We see that we do not only find the decay rate of the probability $\alpha(B)$, but also the equilibrium distribution of the Markov chain that goes with $\xi^{(i)*}$. Note that this optimization is equivalent to executing the following minimization

$$\inf_{\pi:\langle\pi,r\rangle>0}\frac{I_2(\pi)}{\langle\pi,r\rangle} \tag{5}$$

over all probability measures π on E.

This method has several drawbacks. Firstly, in general the function $I_2(\cdot)$ cannot be evaluated easily. Only in very few special cases a simple closed-form expression for $I_2(\cdot)$ can be calculated, for instance in case of birth death Markov chains:

$$I_2(\pi) = \sum_{i=1}^{N}\pi(i)\lambda_{ii} - 2\sum_{i=1}^{N-1}\sqrt{\pi(i)\pi(i+1)\lambda_{i+1,i}\lambda_{i,i+1}}\ ,$$

see also Ridder [1991]. Secondly, it is true that we derive an important characterization of the underlying Markov chain conditioned on an overflow (B large), namely its long-run distribution π_M, but this does not determine uniquely the complete probabilistic behavior of this chain. Therefore, we have to find the transition matrix M under condition of a loss cycle. The knowledge of these rates enables us to realize variance reduction in simulation, as we will see in section 7.

4 A Derivation of the level 3 Rate Function

To cope with the second drawback mentioned in the previous section, we must not only consider the fraction of time the process spends in each state, but also the fraction of jumps from $i \in E$ to $j \in E$, where $i \neq j$. It will appear that this method also removes the first drawback! In fact, this approach can be seen as a level 3 method, because we consider large deviations of the empirical distribution of pairs of subsequent random variables, see Ellis [1985].

To explain the main ideas of this level 3 method we first consider an irreducible discrete time Markov chain $\{X_n, n \in \mathbb{N}_0\}$ on a finite state space, say $E := \{1, \ldots, N\}$. Let $P = (p_{ij})_{i,j=1}^N$ be the matrix containing the transition probabilities; π_P is its uniquely determined long-run distribution. Let $R_n(i,j)$ be the *empirical pair measure*:

$$R_n(i,j) := \frac{1}{n} \sum_{k=1}^{n} 1_{\{i\}}(X_{k-1}) 1_{\{j\}}(X_k).$$

As we know from probability, $R_n(i,j)$ converges to $\pi_P(i)p_{ij}$, almost surely $(n \to \infty)$. Note that for large n it holds that both marginals of $R_n(i,j)$ are identical:

$$\lim_{n\to\infty} \sum_{j=1}^{N} R_n(i,j) = \lim_{n\to\infty} \sum_{j=1}^{N} R_n(j,i)$$

with probability 1. It is clear that, in case of the large deviations of $R_n(\cdot,\cdot)$, we only have to consider matrices $R = (r_{ij})_{i,j=1}^N$ with all entries non-negative, adding up to 1 and $\sum_{j=1}^N r_{ij} = \sum_{j=1}^N r_{ji}$. As can be found in Dembo and Zeitouni [1993], $R_n(\cdot,\cdot)$ obeys a large deviations principle with rate function

$$I_3(R) := \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij} \log\left(\frac{r_{ij}}{p_{ij}}\right) - \sum_{i=1}^{N} r_{i\cdot} \log r_{i\cdot},$$

with $r_{i\cdot} := \sum_{j=1}^N r_{ij}$. Here $0 \log 0$ and $\log(0/0)$ are defined to be 0. Note that $p_{ij} = 0$ and $r_{ij} > 0$ yields $I_3(R) = \infty$.

The statements in the previous paragraph need some explanation. Let Φ_E be the space of probability measures on E. Now define the so-called variational distance between the probability measures μ and ν in Φ_E: $\mathcal{D}(\mu,\nu) := \sup_{A \subseteq E} \mid \nu(A) - \mu(A) \mid$. Having defined this metric, we defined implicitly open and closed sets on Φ_E. The large deviations principle above states that for any closed set F and any open set G in $\Phi_{E\times E}$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathsf{P}(R_n(\cdot,\cdot) \in F) \leq \inf_{R\in F} I_3(R),$$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathsf{P}(R_n(\cdot,\cdot) \in G) \geq \inf_{R\in G} I_3(R).$$

Heuristically writing, we get the following rate of convergence of the probability that the empirical pair measure lies in a small neighborhood of a certain matrix R:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathsf{P}(R_n(\cdot,\cdot) \approx R) = I_3(R).$$

Note that inserting $r_{ij} := \pi_P(i)p_{ij}$ indeed yields $I_3(R) = 0$.

If we found an empirical pair measure R it is very easy to find a transition matrix $Q = (q_{ij})_{i,j=1}^N$ such that $r_{ij} = \pi_Q(i)q_{ij}$: take $q_{ij} := (r_{ij}) \Big/ \left(\sum_{k=1}^N r_{ik}\right)$, $i,j \in \{1,\ldots,N\}$. To prove the correctness of this assertion, it is sufficient to show that $\pi_Q(i) = \sum_{k=1}^N r_{ik}$.

As usual, π_Q denotes the vector of steady state probabilities when transition matrix Q is involved. This can be done as follows:

$$x_i := \sum_{k=1}^{N} r_{ik} \stackrel{!}{=} \sum_{k=1}^{N} r_{ki} = \sum_{k=1}^{N} \left(\sum_{j=1}^{N} r_{kj} \right) q_{ki} = \sum_{k=1}^{N} x_k q_{ki}.$$

Furthermore, from the nature of the matrix R, we have $\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij} = 1$. Invoking the uniqueness of the invariant probability measure, the stated follows directly.

By this relation between empirical pair measures and transition matrices the entropy function $I_3(\cdot)$ can be rewritten as a *relative entropy* between two transition matrices:

$$\begin{aligned} I_3(Q \mid P) &:= \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_Q(i) q_{ij} \log \left(\frac{\pi_Q(i) q_{ij}}{p_{ij}} \right) - \sum_{i=1}^{N} \pi_Q(i) \log \pi_Q(i) \\ &= \sum_{i=1}^{N} \pi_Q(i) \sum_{j=1}^{N} q_{ij} \log \left(\frac{q_{ij}}{p_{ij}} \right). \end{aligned}$$

Since for each state i the function $\sum_{j=1}^{N} q_{ij} \log(q_{ij}/p_{ij})$ equals 0 for $p_{ij} = q_{ij}$ (all j) and is positive elsewhere, we see that $I_3(Q \mid P)$ is non-negative and equal 0 if and only if $P = Q$. The notion 'relative entropy' enables us to examine the probability that a Markov chain with actual transition matrix P acts as if obeying different transition probabilities (the entries of Q) during a substantial period. In fact, $I_3(Q \mid P)$ provides us the exponent of the probability of observing a chain that behaves according to Q instead of P.

Now we derive also a relative entropy with respect to a transition matrix $\Lambda := (\lambda_{ij})_{i,j=1}^{N}$ of a continuous time Markov chain $X(\cdot)$. Let $\{X_n^\epsilon, n \in \mathbb{N}_0\}$ be the ϵ-discretized version of $X(\cdot)$: $X_n^\epsilon := X(n\epsilon)$, with transition probabilities

$$p_{ij}^\epsilon := \mathsf{P}\big(X(n\epsilon + \epsilon) = j \mid X(n\epsilon) = i\big), \text{ where } \lambda_{ij} = \lim_{\epsilon \downarrow 0} \frac{p_{ij}^\epsilon}{\epsilon} \ (j \neq i) \ , \ \lambda_{ii} = \lim_{\epsilon \downarrow 0} \frac{p_{ii}^\epsilon - 1}{\epsilon}.$$

For another continuous time Markov chain transition matrix $M = (\mu_{ij})_{i,j=1}^{N}$, we find q_{ij}^ϵ in the same way. Following the same line of reasoning as in Donsker and Varadhan [1975], we find results for continuous time Markov chains from those for discrete time by means of discretization. Heuristically, we deduce

$$\begin{aligned} &\lim_{t \to \infty} \frac{1}{t} \log \mathsf{P}\big(X(\cdot) \text{ behaves as under } M \text{ instead of } \Lambda \text{ on } [0,t]\big) \\ &= \lim_{t \to \infty} \frac{1}{t} \log \lim_{\epsilon \downarrow 0} \mathsf{P}\big(\{X_k^\epsilon\} \text{ behaves as under } Q^\epsilon \text{ instead of } P^\epsilon \text{ on } \{0, \ldots, \lfloor t/\epsilon \rfloor\}\big) \\ &= \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \lim_{t \to \infty} \frac{1}{\lfloor t/\epsilon \rfloor} \log \mathsf{P}\big(\{X_k^\epsilon\} \text{ behaves as under } Q^\epsilon \text{ instead of } P^\epsilon \text{ on } \{0, \ldots, \lfloor t/\epsilon \rfloor\}\big) \\ &= -\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \sum_{i=1}^{N} \pi_{Q^\epsilon}(i) \sum_{j=1}^{N} q_{ij}^\epsilon \log \left(\frac{q_{ij}^\epsilon}{p_{ij}^\epsilon} \right). \end{aligned}$$

It is easy to derive that $\pi_{Q^\epsilon} \to \pi_M$, where $\epsilon \downarrow 0$ and π_M denotes the steady state distribution of a continuous time Markov chain with transition matrix M. For distinct i and j we find that the ratio of transition probabilities p_{ij}^ϵ and q_{ij}^ϵ approaches the ratio of transition rates λ_{ij} and μ_{ij}, where ϵ approaches 0:

$$\log\left(\frac{q_{ij}^\epsilon}{p_{ij}^\epsilon}\right) = \log\left(\frac{q_{ij}^\epsilon/\epsilon}{p_{ij}^\epsilon/\epsilon}\right) \to \log\left(\frac{\mu_{ij}}{\lambda_{ij}}\right), \quad \epsilon \downarrow 0.$$

Since $p_{ij}^\epsilon \to 0$ if $\epsilon \downarrow 0$, we find using Taylor expansions,

$$\lim_{\epsilon \downarrow 0} \frac{\log p_{ii}^\epsilon}{\epsilon} = \lim_{\epsilon \downarrow 0} \frac{\log(1 - \sum_{j \neq i} p_{ij}^\epsilon)}{\epsilon} = -\lim_{\epsilon \downarrow 0} \frac{\sum_{j \neq i} p_{ij}^\epsilon}{\epsilon} = -\sum_{j \neq i} \lambda_{ij} = \lambda_{ii},$$

and an analogous expression for $\log q_{ii}^\epsilon/\epsilon$. We find for the relative entropy of M with respect to Λ:

$$I_3(M \mid \Lambda) := \sum_{i=1}^N \pi_M(i) \sum_{j=1, j \neq i}^N \mu_{ij} \log\left(\frac{\mu_{ij}}{\lambda_{ij}}\right) + \sum_{i=1}^N \pi_M(i)\big(\mu_{ii} - \lambda_{ii}\big),$$

which can also be found in Kesidis and Walrand [to appear]. One may show that the following contraction principle holds:

$$I_2(\pi_M) = \inf_{M : \pi_M M = 0} I_3(M \mid \Lambda),$$

cf. De Veciana *et al.* [1993] and equation (3.1.18) in Dembo and Zeitouni [1993]. Therefore, the optimization programs

$$\inf_{\pi_M : \langle \pi_M, r \rangle > 0} \frac{I_2(\pi_M)}{\langle \pi_M, r \rangle} \quad \text{and} \quad \inf_{M :\, \pi_M M = 0, \langle \pi_M, r \rangle > 0} \frac{I_3(M \mid \Lambda)}{\langle \pi_M, r \rangle} \tag{6}$$

are equivalent.

5 Necessary conditions for the optimizing transition matrix

THIS SECTION FOCUSES on the optimization program

$$\inf_{\widetilde{M} : \langle \pi_{\widetilde{M}}, r \rangle > 0} \frac{I_3(\widetilde{M} \mid \Lambda)}{\langle \pi_{\widetilde{M}}, r \rangle}, \tag{7}$$

$\pi_{\widetilde{M}}$ defined as the long-run distribution of the continuous time Markov chain with transition matrix $\widetilde{M}$. Because of equation (6), the infimum above equals (5), and therefore also θ^*. This minimization is to be preferred to the level 2 method, because the large deviation rate function $I_3(\cdot \mid \Lambda)$ can be evaluated a lot easier than $I_2(\cdot)$. Apart from that, we may derive all transition rates (and not only the equilibrium distribution) of the chain under

condition of reaching overflow. In this section we try to find some characteristics for the optimizing transition matrix, say M.

In section 2 it was explained how to find an associated random variable $\xi^{(i)*}$ of $\xi^{(i)}$ satisfying the translation relation (4). Clearly we have $\mathsf{E}\exp(-\theta^*\xi^{(i)*}) = 1$, yielding that $\theta^{**} = -\theta^*$. Again applying (4), we get that the associated of this associated is again the original random variable: $\xi^{(i)**} = \xi^{(i)}$. Moreover, we find easily

$$\lim_{B\to\infty}\frac{1}{B}\log \mathsf{P}\left(\left\{\sum_{k=1}^{j}\xi_k^{(i)}, j\in\mathbb{N}\right\} \text{ hits } [B,\infty)\right) =$$

$$\lim_{B\to\infty}\frac{1}{B}\log \mathsf{P}\left(\left\{-\sum_{k=1}^{j}\xi_k^{(i)*}, j\in\mathbb{N}\right\} \text{ hits } [B,\infty)\right) = -\theta^*.$$

We may deduce similarly to lemma 1 in Ridder [1993], that a transition matrix that goes with $\xi^{(i)*}$ equals the minimizing argument in the optimization program (7), which we called M.

Combining the observations of the previous paragraph, we have the following duality relation:

$$\frac{I_3(M \mid \Lambda)}{\langle \pi_M, r\rangle} = \inf_{\widetilde{M}:\langle \pi_{\widetilde{M}}, r\rangle > 0} \frac{I_3(\widetilde{M} \mid \Lambda)}{\langle \pi_{\widetilde{M}}, r\rangle} = \theta^* = \inf_{\widetilde{\Lambda}:\langle \pi_{\widetilde{\Lambda}}, r\rangle < 0} \frac{I_3(\widetilde{\Lambda} \mid M)}{-\langle \pi_{\widetilde{\Lambda}}, r\rangle} = \frac{I_3(\Lambda \mid M)}{-\langle \pi_\Lambda, r\rangle}.$$

From the definition of the relative entropy $I_3(\cdot \mid \cdot)$, we conclude that the fact that θ^* is finite implies that $\lambda_{ij} = 0 \Longleftrightarrow \mu_{ij} = 0$ (for $i, j \in E, i \neq j$). We derived that the Markov chains equipped with transition matrices Λ and M, respectively, have the same chain structure. We conclude that the chain governed by M is irreducible as well; given M, π_M is uniquely determined and consists of merely positive components.

In order to solve (7), we introduce the *Lagrangian* which has to be minimized over the transition matrices $\widetilde{M}$ with $\langle \pi_{\widetilde{M}}, r\rangle > 0$ and which contains already the condition $\pi_{\widetilde{M}}\widetilde{M} = 0$:

$$L(\widetilde{M}, \pi_{\widetilde{M}}, \widetilde{K}) := \frac{I_3(\widetilde{M} \mid \Lambda)}{\langle \pi_{\widetilde{M}}, r\rangle} - \sum_{i=1}^{N}\widetilde{K}_i\left(\sum_{j\neq i}\widetilde{\mu}_{ij}\pi_{\widetilde{M}}(i) - \sum_{j\neq i}\widetilde{\mu}_{ji}\pi_{\widetilde{M}}(j)\right)$$

$$= \frac{\sum_{i=1}^{N}\left(\pi_{\widetilde{M}}(i)\sum_{j=1, j\neq i}^{N} f(i,j)\right)}{\langle \pi_{\widetilde{M}}, r\rangle} - \sum_{i=1}^{N}\widetilde{K}_i\, g(i),$$

where $f(\cdot,\cdot)$ and $g(\cdot)$ are defined for fixed $\widetilde{M}$ and $\pi_{\widetilde{M}}$:

$$\begin{aligned} f(i,j) &:= \widetilde{\mu}_{ij}\log(\widetilde{\mu}_{ij}/\lambda_{ij}) + \lambda_{ij} - \widetilde{\mu}_{ij} \text{ for distinct } i \text{ and } j \in E \\ g(i) &:= \sum_{j\neq i}\widetilde{\mu}_{ij}\pi_{\widetilde{M}}(i) - \sum_{j\neq i}\widetilde{\mu}_{ji}\pi_{\widetilde{M}}(j) \text{ for } i \in E. \end{aligned}$$

We make a couple of remarks:

- Since $I_3(\cdot \mid \Lambda)$ as well as $\langle \cdot, r \rangle$ are proportional in $\pi_{\widetilde{M}}$, we may omit the normalization equation $\sum_{i=1}^{N} \pi_{\widetilde{M}}(i) = 1$.

- For notational convenience, we did not remove the equation $g(N) = 0$, although redundant.

It is a well-known result that a necessary condition for the minimum is that all partial derivatives equal 0. Equating the partial derivatives of the $\widetilde{\mu}_{ij}$ (where $i \neq j$), $\pi_{\widetilde{M}}(i)$ and $\widetilde{K}_i$ to zero yields subsequently:

$$\frac{\partial L}{\partial \widetilde{\mu}_{ij}} = \pi_{\widetilde{M}}(i) \left(\frac{\log(\widetilde{\mu}_{ij}/\lambda_{ij})}{\langle \pi_{\widetilde{M}}, r \rangle} - \widetilde{K}_i + \widetilde{K}_j \right) = 0 ; \tag{8}$$

$$\begin{aligned} \frac{\partial L}{\partial \pi_{\widetilde{M}}(i)} &= \frac{\sum_{j \neq i} f(i,j) \langle \pi_{\widetilde{M}}, r \rangle - \left(\sum_{k=1}^{N} \pi_{\widetilde{M}}(k) \sum_{j \neq k} f(k,j) \right) r(i)}{\langle \pi_{\widetilde{M}}, r \rangle^2} - \widetilde{K}_i \sum_{j \neq i} \widetilde{\mu}_{ij} + \sum_{j \neq i} \widetilde{K}_j \widetilde{\mu}_{ij} \\ &= \frac{\sum_{j \neq i} f(i,j)}{\langle \pi_{\widetilde{M}}, r \rangle} - \frac{\left(\sum_{k=1}^{N} \pi_{\widetilde{M}}(k) \sum_{j \neq k} f(k,j) \right) r(i)}{\langle \pi_{\widetilde{M}}, r \rangle^2} + \sum_{j \neq i} (\widetilde{K}_j - \widetilde{K}_i) \widetilde{\mu}_{ij} = 0 ; \end{aligned} \tag{9}$$

$$\frac{\partial L}{\partial \widetilde{K}_i} = - \sum_{j \neq i} \widetilde{\mu}_{ij} \pi_{\widetilde{M}}(i) + \sum_{j \neq i} \widetilde{\mu}_{ji} \pi_{\widetilde{M}}(j) = 0 \ \big(\text{or } g(i) = 0 \big) . \tag{10}$$

From now on we examine relations (8), (9) and (10) a bit more precisely. Suppose that the infimum is attained in (M, π_M, K). Since the vector π_M consists of only positive entries, equation (8) yields $\log(\mu_{ij}/\lambda_{ij}) = \langle \pi_M, r \rangle (K_i - K_j)$, where $i, j \in E$ and $i \neq j$. Now consider an arbitrary subset $\{i_0, i_1, \ldots, i_n\}$ of E, with $n \in \mathbb{N}$ and satisfying the following constraints: $i_j \neq i_{j+1}$, $j = 0, \ldots, n-1$, $i_0 \neq i_n$. It follows immediately that (defining $i_{n+1} := i_0$)

$$\log \left(\frac{\mu_{i_0 i_1}}{\lambda_{i_0 i_1}} \right) + \cdots + \log \left(\frac{\mu_{i_n i_0}}{\lambda_{i_n i_0}} \right) = \langle \pi_M, r \rangle \times \left(\sum_{j=1}^{n+1} (K_{i_{j-1}} - K_{i_j}) \right) .$$

Clearly, the right-hand side of the previous display equals 0. We see that we obtain the following necessary condition:

- for any subset $\{i_0, i_1, \ldots, i_n\}$ of the state space E (with $n \in \mathbb{N}$ and satisfying $i_j \neq i_{j+1}$, $j = 0, \ldots, n-1$, $i_0 \neq i_n$)

$$\lambda_{i_0 i_1} \cdots \lambda_{i_n i_0} = \mu_{i_0 i_1} \cdots \mu_{i_n i_0} . \tag{11}$$

  In words: the products of subsequent transition rates in cycles of the chain of an optimizing matrix M are equal to these products of the original chain with transition matrix Λ.

Some elementary calculus yields for all $i \in E$

$$\frac{L(\widetilde{M}, \pi_{\widetilde{M}}, \widetilde{K})\, r(i)}{\langle \pi_{\widetilde{M}}, r\rangle} = -\frac{\partial L}{\partial \pi_{\widetilde{M}}(i)} + \left(\sum_{j\neq i} \frac{f(i,j)}{\langle \pi_{\widetilde{M}}, r\rangle} + \sum_{j\neq i} (\widetilde{K}_j - \widetilde{K}_i)\widetilde{\mu}_{ij}\right) - \frac{r(i)}{\langle \pi_{\widetilde{M}}, r\rangle} \sum_{j=1}^{N} \widetilde{K}_j\, g(j).$$

Under the necessary conditions (9) and (10), the first and the last term of the right-hand side equal 0 in (M, π_M, K). Since θ^* is the value of the Lagrangian in a certain stationary point, plugging in condition (8), $\theta^* r(i)/\langle \pi_M, r\rangle$ can be rewritten as follows:

$$\begin{aligned}\frac{\theta^* r(i)}{\langle \pi_M, r\rangle} &= \left(\sum_{j\neq i} \frac{f(i,j)}{\langle \pi_M, r\rangle} + \sum_{j\neq i} (K_j - K_i)\mu_{ij}\right) \\ &= \left(\sum_{j\neq i} \frac{f(i,j)}{\langle \pi_M, r\rangle} + \sum_{j\neq i} \frac{\mu_{ij}\log(\mu_{ij}/\lambda_{ij})}{\langle \pi_M, r\rangle}\right) = \sum_{j\neq i} \frac{(\lambda_{ij} - \mu_{ij})}{\langle \pi_M, r\rangle}.\end{aligned}$$

We found this second necessary condition:

- for any $i \in E$, we have

$$\mu_i = \lambda_i - \theta^* r(i). \tag{12}$$

We found a simple relation between the rowsums of the original transition matrix Λ and those of a minimizing matrix M. We see that in case of $r(i) < 0$ for some state i, we have that $\mu_i > \lambda_i$. This means that average time the chain spends per visit in state i is under M smaller than under Λ. Conversely, if $r(i) > 0$, the mean time of a visit to state i is shorter under Λ than under M. This is intuitively reasonable: if the fluid process reaches a high level the visits to states with a positive rate must be longer and those to states with a negative rate shorter than in the original chain.

Note that the condition $\langle \pi_M, r\rangle > 0$ excludes the trivial solution $M = \Lambda$ and $\theta^* = 0$.

6 Sufficient conditions for the optimizing transition matrix

IN THE PREVIOUS section necessary conditions for an associated transition matrix M of Λ were presented. Of course, the question arises whether the conditions are also sufficient.

Now suppose that there are positive μ_{ij} $(i \neq j)$ and a positive θ^* satisfying both necessary conditions (11) and (12). Consider an arbitrary cycle from state 1 to state 1, visiting

$$1 = i_0 \to i_1 \to \cdots \to i_n \to i_0 = 1,$$

satisfying $i_j \neq i_{j+1}$, $j = 0, \ldots, n-1$, $i_0 \neq i_n$. The chain stays a time t_k in state i_k. The probability of this realization to occur is

$$\left(\frac{\lambda_{i_0 i_1}}{\lambda_{i_0}}\right) \cdots \left(\frac{\lambda_{i_n i_0}}{\lambda_{i_n}}\right) \times \left(\lambda_{i_0} \exp(-\lambda_{i_0} t_0)\right) \cdots \left(\lambda_{i_n} \exp(-\lambda_{i_n} t_n)\right) \mathrm{d}t_0 \cdots \mathrm{d}t_n. \tag{13}$$

So the moment generating function of $\xi^{(1)}$ in $\theta+\theta^*$ equals

$$\sum_{\text{all cycles}} \int_{-\infty}^{\infty} \int_{t_j:\sum r(i_j)t_j=x} e^{(\theta+\theta^*)x}\left(\lambda_{i_0 i_1}\cdots\lambda_{i_n i_0}\right) \times \left(e^{-\lambda_{i_0}t_0}\cdots e^{-\lambda_{i_n}t_n}\right)\mathrm{d}t_0\cdots\mathrm{d}t_n\mathrm{d}x,$$

where the summation is over all cycles starting and ending in 1 and the inner integration over all t_j satisfying $\sum_{j=0}^{n} r(i_j)t_j = x$. Since the μ_{ij} and θ^* satisfy the two conditions (11) and (12), we conclude that the expression above is equal to

$$\sum_{\text{all cycles}} \int_{-\infty}^{\infty} \int_{t_j:\sum r(i_j)t_j=x} e^{\theta x}\left(\mu_{i_0 i_1}\cdots\mu_{i_n i_0}\right) \times \left(e^{-\mu_{i_0}t_0}\cdots e^{-\mu_{i_n}t_n}\right)\mathrm{d}t_0\cdots\mathrm{d}t_n\mathrm{d}x,$$

cf. Ridder [1993]. We see that for these particular M and θ^* equation (4) is satisfied! Recalling that the transition matrix associated with $\xi^{(1)*}$ equals the minimizer of (7), we conclude that both conditions (11) and (12) are necessary as well as sufficient. Remark that, because both conditions are invariant under the choice of the initial state i, we found that the same Markov chain goes also with $\xi^{(i)*}$, for $i \neq 1$.

In case of a full matrix Λ we will show that the equations in the two conditions consist of an equal number of equations and variables, considering θ^* as unknown, namely N^2-N+1 (i.e., the number of entries in the transition matrix M as well as the decay rate of $\alpha(B)$, namely θ^*). We notice that condition (12) consists of N equations. Therefore, we now focus on condition (11).

- Obviously there are $\binom{N}{2} = \frac{1}{2}N(N-1)$ independent equations concerning two different states in E.

- From some simple combinatorical arguments it can be seen that we do not need all $\binom{N}{3}$ equations with three different states, but only $\frac{1}{2}(N-1)(N-2)$. This can be shown as follows. Consider all sets of three distinct states containing state 1. For each of the $\binom{N-1}{2}$ sets $\{1,i,j\}$ choose a direction $1 \to i \to j \to 1$. Suppose that $\lambda_{1i}\lambda_{ij}\lambda_{j1} = \mu_{1i}\mu_{ij}\mu_{j1}$, then also

$$\lambda_{1j}\lambda_{ji}\lambda_{i1} = \frac{(\lambda_{1j}\lambda_{j1})(\lambda_{ji}\lambda_{ij})(\lambda_{i1}\lambda_{1i})}{\lambda_{1i}\lambda_{ij}\lambda_{j1}} = \frac{(\mu_{1j}\mu_{j1})(\mu_{ji}\mu_{ij})(\mu_{i1}\mu_{1i})}{\mu_{1i}\mu_{ij}\mu_{j1}} = \mu_{1j}\mu_{ji}\mu_{i1}.$$

  So condition (11) applies to all sets of three distinct states containing 1. Now consider a set $\{i,j,k\}$ not containing 1. In a similar way

$$\lambda_{ij}\lambda_{jk}\lambda_{ki} = \frac{(\lambda_{ij}\lambda_{j1}\lambda_{1i})(\lambda_{jk}\lambda_{k1}\lambda_{1j})(\lambda_{ki}\lambda_{i1}\lambda_{1k})}{(\lambda_{j1}\lambda_{1j})(\lambda_{i1}\lambda_{1i})(\lambda_{k1}\lambda_{1k})}$$

$$= \frac{(\mu_{ij}\mu_{j1}\mu_{1i})(\mu_{jk}\mu_{k1}\mu_{1j})(\mu_{ki}\mu_{i1}\mu_{1k})}{(\mu_{j1}\mu_{1j})(\mu_{i1}\mu_{1i})(\mu_{k1}\mu_{1k})} = \mu_{ij}\mu_{jk}\mu_{ki}.$$

- It can be shown quite easily that, once having found μ_{ij} satisfying for sets containing two or three states, they also satisfy for all sets of n distinct states, $n \geq 4$. Take for instance $i \to j \to k \to l \to i$. Then obviously

$$\begin{aligned} \lambda_{ij}\lambda_{jk}\lambda_{kl}\lambda_{li} &= \frac{(\lambda_{ij}\lambda_{jk}\lambda_{ki})(\lambda_{ik}\lambda_{kl}\lambda_{li})}{\lambda_{ki}\lambda_{ik}} \\ &= \frac{(\mu_{ij}\mu_{jk}\mu_{ki})(\mu_{ik}\mu_{kl}\mu_{li})}{\mu_{ki}\mu_{ik}} = \mu_{ij}\mu_{jk}\mu_{kl}\mu_{li}. \end{aligned}$$

  This trick applies to all sets of n distinct states, $n \geq 4$.

We found that there are $N + \frac{1}{2}N(N-1) + \frac{1}{2}(N-1)(N-2) = N^2 - N + 1$ equations in the same number of unknowns. If Λ contains some zeros, we already saw that M has zeros on the same positions. Provided that Λ is ergodic, one may derive also in this case the equality between the number of equalities and unknowns. We see immediately that $\theta^* = 0$ and $\mu_{ij} = \lambda_{ij}$ is a solution to the equations, but this trivial solution corresponds to the root 0 of $\mathsf{E}\exp(\theta\xi^{(i)}) = 1$.

Having found a system with the same number of equations and unknowns, we may try to solve it numerically, using Newton-like methods.

A remarkable result is the following. Consider a fluid model modulated by a birth death continuous time Markov chain. Once having computed θ^* in the simple way described in section 2, we find μ_{12} using condition (12); this yields μ_{21} with rule (11); applying (12) again we obtain μ_{23}, etc. We conclude that we found a nice recursive method to compute M.

De Veciana *et al.* [1993] proposed a method to find θ^* and M by simply solving the system of equations mentioned in this section, in this case $2N - 1$ equations in $2N - 1$ unknowns.

- If we are only interested in the 'exponent of $\alpha(B)$', our approach described in section 2 is probably easier to implement. Apart from that, the program of De Veciana *et al.* [1993] only provides a solution in the birth death case and not in general.

- If we do not want to obtain only θ^*, but also M (in the birth death case), the method mentioned in section 2 (to compute θ^*) in conjunction with the recursion scheme described above will probably be easier to perform than the system proposed by De Veciana *et al* [1993].

7 Fast Simulation by Importance Sampling

IN FINITE BUFFER models the choice of buffersize B is often an important issue. We may choose B for instance such that the probability of an arbitrary cycle being a loss cycle,

$\alpha(B)$, is very small, typically in the order of 10^{-9}. For these design purposes it is useful to have good estimates of $\alpha(B)$.

From the preceding sections we found the decay rate of $\alpha(B)$, but not an approximation of the probability itself. Therefore, we may try to estimate it by means of *Monte Carlo simulation*: generate cycles of the fluid process and estimate $\alpha(B)$ by the ratio of the number of overflow cycles to the total number of cycles. For large levels B the number of samples to draw must be large in order to obtain good relative efficiency of the estimate. As can be found in Walrand [1988], in case of confidence b and efficiency - i.e., relative width of the confidence interval - a, the number of cycles that should be drawn is approximately $n_b^2/\alpha(B)a^2$, where n_b denotes the $\frac{1}{2}+\frac{b}{2}$-percentile of the standard Normal distribution. E.g., when b should be 95% and a 10%, this number is about $400/\alpha(B)$. Is $\alpha(B)$ of the order 10^{-9}, then the number of cycles required is of the order 10^{11}, which imposes strong demands on the random generator. Apart from that, this Monte Carlo method would obviously be very time consuming.

To cope with the problems mentioned in the previous paragraph, we may tackle the problem with a technique called *Importance Sampling*. For some basic notes on this subject we refer to Glynn and Iglehart [1989] and Siegmund [1976].

We assume some underlying probability triple $(\Omega, \mathcal{F}, \mathsf{P})$, under which the original Monte Carlo simulation would be executed. We may specify the sample space Ω as follows. Each sample $\omega \in \Omega$ represents either the complete busy period of a regular cycle in the buffer process - no overflow - or the first part of an overflow cycle in which the buffer process builds up from 0 to level B. Each ω is of the form

$$\omega = \big((i_0,t_0),\ (i_1,t_1),\ \ldots,\ (i_n,t_n)\big),$$

where i_k represents the state of the chain after the kth jump during this sample cycle and t_k measures the amount of time spent in state i_k. Notice that $\sum_{k=0}^{m} r(i_k)t_k$ represents the buffer occupancy just before the $(m+1)$th jump. Clearly a cycle is a loss cycle if $\sum_{k=0}^{n} r(i_k)t_k \geq B$, whereas it represents a regular cycle if $\sum_{k=0}^{n} r(i_k)t_k \leq 0$. Consider then the possibility of executing the simulation based on another probability Q, such that P is absolutely continuous relative to Q. Let $L(\cdot)$ be a *Radon-Nikodým derivative* or likelihood function:

$$L(\omega) := \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}}(\omega).$$

We may estimate the desired probability by the sum of the likelihood of the cycles that reached B divided by the total number of cycles. It is a well-known result that under the sufficient condition that these likelihoods in overflow cycles are smaller than 1, the resulting estimator has a smaller variance than the direct Monte Carlo estimator (see Walrand [1988]). As a consequence, it suffices to generate substantially less cycles under Q than under P.

Since the chain behaves in case of buffer overflows as under M instead of Λ it seems reasonable to associate Q with the fluid model governed by an underlying Markov chain with rates M. Then the likelihood of a cycle satisfies

$$L(\omega) = \frac{\lambda_{i_0 i_1} \cdots \lambda_{i_{n-1} i_n} \lambda_{i_n}}{\mu_{i_0 i_1} \cdots \mu_{i_{n-1} i_n} \mu_{i_n}} \exp\left(-\theta^* \sum_{k=0}^{n} r(i_k) t_k\right),$$

cf. density (13). In the cycles reaching B, the exponent involved in the previous likelihood is smaller than $\exp(-\theta^* B)$. Bearing in mind that condition (11) holds and that B is large, this likelihood will be smaller than 1 in most of the cases.

It is a well-known result that in case of an overflow the buffer contents builds up essentially linear with slope $\langle \pi_M, r \rangle > 0$ (cf. Anantharam [1988]), exactly the drift of the fluid process under M. We find that the average behavior of the fluid system under M looks like the behavior under Λ under condition of an overflow. The drift of the new process becomes positive instead of $\langle \pi_\Lambda, r \rangle < 0$.

The main ideas behind this 'fast simulation' in conjunction with large deviations theory can be found in Cottrell *et al.* [1983] and Parekh and Walrand [1989]. In the former it is shown that the described change of measures is in some sense optimal within the class of exponential twisted distributions, in the latter this result is applied to rare events in a queueing context.

8 Simulation results

FOLLOWING THE PROCEDURES explained in the previous section, we now work out the concepts of 'fast simulation'. It appears that the 'fast simulation' method provides us a possibility to speed up simulation runs substantially. The rates of the dual transition matrices can be calculated from the optimization program (7) or the system of equations proposed in section 6.

• We first consider a Markov chain on $\{1, \ldots, 5\}$ with rates $r(1) = 1$ and $r(i) = -1$ for $i = 2, \ldots, 5$. This means that state 1 represents the bursty moments and the other states the quiet moments during a communication connection.

$$\Lambda := \begin{pmatrix} * & 2 & 1 & 1 & 1 \\ 2 & * & 3 & 2 & 1 \\ 4 & 2 & * & 1 & 1 \\ 1 & 1 & 1 & * & 2 \\ 1 & 3 & 1 & 1 & * \end{pmatrix}, \qquad M = \begin{pmatrix} * & 0.807 & 0.494 & 0.319 & 0.334 \\ 4.959 & * & 3.677 & 1.582 & 0.829 \\ 8.093 & 1.632 & * & 0.645 & 0.676 \\ 3.136 & 1.265 & 1.550 & * & 2.096 \\ 2.992 & 3.620 & 1.479 & 0.954 & * \end{pmatrix}.$$

Note that a cycle necessarily starts in 1, since only $r(1) > 0$. We remark that in this case the likelihood in a loss cycle may be simplified to

$$L(\omega) = \frac{\lambda_1}{\mu_1} \exp\left(-\theta^* \sum_{i=0}^{n} r(i_k) t_k\right).$$

The equilibrium distributions of Λ and M are

$$\pi_\Lambda = \begin{pmatrix} 0.277 \\ 0.197 \\ 0.155 \\ 0.200 \\ 0.171 \end{pmatrix}, \quad \pi_M = \begin{pmatrix} 0.722 \\ 0.0885 \\ 0.0772 \\ 0.0586 \\ 0.0541 \end{pmatrix}, \quad \begin{aligned} \langle \pi_\Lambda, r \rangle &= -0.446, \\ \langle \pi_M, r \rangle &= 0.443. \end{aligned}$$

with $\theta^* = 3.046$. We see that we indeed changed the negative drift into a positive one.

In the next table, we first present simulation results using both simulation methods mentioned in the previous section: 'fast simulation' using transition matrix M and the direct Monte Carlo method.

Table 1: Simulation results Example 1.

B	2.5		3.0		7.0	
	Fast	Direct	Fast	Direct	Fast	LD
# cycles	10^3	10^6	10^3	10^6	10^3	
$\widehat{\alpha(B)}$	$2.96 \cdot 10^{-4}$	$3.01 \cdot 10^{-4}$	$6.32 \cdot 10^{-5}$	$5.60 \cdot 10^{-5}$	$3.44 \cdot 10^{-10}$	
RE	7.9%	11.3%	8.0%	26.2%	7.8%	
$\widehat{\theta^*}$	3.25	3.24	3.22	3.26	3.11	3.046
$\widehat{T(B)}$	5.14	5.06	6.30	5.93	15.33	
RE	4.9%	5.2%	3.7%	8.6%	2.1%	
$\widehat{T(B)}/B$	2.06	2.02	2.10	1.98	2.19	2.257
Time	4:23	7:16:27	4:84	8:23:45	5:06	

We first estimated the probability of a loss cycle. Then we gave the relative efficiency of these estimates: the ratio of the confidence interval half length to the estimated value. We fixed the confidence on 95%. As we saw before $-\log\big(\alpha(B)\big)/B$ approaches θ^* for large B. With $\widehat{\theta^*}$ we denote $-\log\big(\widehat{\alpha(B)}\big)/B$.

We also gave an estimate for the expected time to an overflow in an overflow cycle. As stated in Ridder [1993], this quantity will be asymptotically equal to $B/\langle \pi_M, r \rangle$. Finally, simulation time is given (in minutes). The simulation jobs are performed on a 486 personal computer, IBM 433DXS.

It appears that the estimated values $-\log\big(\widehat{\alpha(B)}\big)/B$ and $\widehat{T(B)}/B$ move only very slowly to the large deviations expressions mentioned in the final column. For larger values of B there is a smaller discrepancy between these values, but in that case the performance of direct simulation will take very much time. Therefore, we omit direct simulation results in the case of $B = 7.0$.

• Now we investigate a chain obeying some transition matrix containing several zero rates, in contrast to the first example. Apart from that, there are multiple states representing the bursty moments.

The Markov chain is defined on the same state space $\{1,\ldots,5\}$ with traffic rate function $r = (1,1,-2,-1,-1)$. In the dual transition matrix M of Λ, the entries corresponding with a zero entry in Λ equal 0.

$$\Lambda := \begin{pmatrix} * & 2 & 0 & 1 & 1 \\ 2 & * & 3 & 0 & 0 \\ 4 & 0 & * & 1 & 0 \\ 0 & 0 & 1 & * & 2 \\ 1 & 3 & 1 & 1 & * \end{pmatrix}, \qquad M = \begin{pmatrix} * & 1.933 & 0 & 0.648 & 0.800 \\ 2.069 & * & 2.312 & 0 & 0 \\ 5.369 & 0 & * & 0.869 & 0 \\ 0 & 0 & 1.150 & * & 2.469 \\ 1.251 & 3.627 & 0.932 & 0.810 & * \end{pmatrix}.$$

We see that under M the chain stays on average a larger fraction of time in the states i with $r(i) > 0$:

$$\pi_\Lambda = \begin{pmatrix} 0.306 \\ 0.194 \\ 0.180 \\ 0.202 \\ 0.118 \end{pmatrix}, \qquad \pi_M = \begin{pmatrix} 0.356 \\ 0.258 \\ 0.0962 \\ 0.167 \\ 0.122 \end{pmatrix}, \qquad \begin{aligned} \langle \pi_\Lambda, r \rangle &= -0.180, \\ \langle \pi_M, r \rangle &= 0.132. \end{aligned}$$

The exponent of the likelihood of observing an overflow is $\theta^* = 0.619$. Since there are multiple states having a positive rate, we let $E^+ \subset E$ contain all states i for which $r(i) > 0$. Suppose that at time 0 the buffer is empty and that the chain starts off in a state $i \in E^+$. The cycle that is originated will be an overflow cycle with probability, say, $\alpha_i(B)$. We let

$$\nu_\Lambda(i) := \frac{\pi_\Lambda(i)}{\sum_{j \in E^+} \pi_\Lambda(j)}$$

be the quasi-stationary distribution. Then our overflow probability is given by

$$\alpha(B) = \sum_{i \in E^+} \nu_\Lambda(i) \alpha_i(B).$$

The simulation results are summarized in the following table:

Table 2: Simulation results Example 2.

B	10.0		15.0		25.0	
	Fast	Direct	Fast	Direct	Fast	LD
# cycles	10^3	10^6	10^3	10^6	10^3	
$\widehat{\alpha(B)}$	$5.22 \cdot 10^{-4}$	$5.54 \cdot 10^{-4}$	$2.65 \cdot 10^{-5}$	$2.10 \cdot 10^{-5}$	$4.61 \cdot 10^{-8}$	
RE	10.6%	8.3%	9.9%	42.8%	11.0%	
$\widehat{\theta^*}$	0.76	0.75	0.70	0.72	0.68	0.619
$\widehat{T(B)}$	43.03	44.97	72.70	76.38	131.30	
RE	5.6%	6.5%	4.7%	15.1%	4.3%	
$\widehat{T(B)}/B$	4.30	4.50	4.85	5.09	5.25	7.578
Time	5:88	16:17:41	6.06	16:58:67	8:63	

We see that the large deviations expressions are approached very slowly. This may be due to the fact that the associated process has only a tiny positive drift.

• We conclude this section with the analysis of fluid models with *multiple independent Markov fluid sources.* In section 1 we associated a customer who is connected to a communication network and who loads packets into a buffer of finite capacity B, with a continuous time Markov chain $X(\cdot)$ and traffic rate function $r = \tilde{r} - c$. The chain describes the time behavior of the connection and the rate function reflects the loading characteristics. In this example we allow several connections loading the same buffer, independently of each other. Suppose that there are K customers connected, then customer k is recorded by a Markov chain $X_k(\cdot)$ on state space E_k with N_k states, equipped with transition matrix $\Lambda^{(k)}$ and rate function $\tilde{r}_k(\cdot)$. It is a matter of simple transformations to obtain the model described in the introduction of this paper. The (vector) process $X(\cdot)$ defined by

$$X(t) := \big(X_1(t), \ldots, X_K(t)\big)$$

is a Markov chain on state space $E_1 \times \cdots \times E_K$ with traffic rate

$$\sum_{k=1}^{K} \tilde{r}_k(i_k) - c \text{ at time } t \text{ if } X_1(t) = i_1, \ldots, X_K(t) = i_K.$$

The new transition matrix, say Λ, is sparse, since the probability of two or more transitions within a small time interval of length Δt is $o(\Delta t)$. If i and j are elements of the new state space, the associated infinitesimal transition rate λ_{ij} may not equal 0 only if i and j differ in exactly one of the K components. Based on the decoupling result established in De Veciana *et al.* [1993], the optimizing transition matrix M can be partitioned in transition matrices $M^{(1)}, \ldots, M^{(K)}$.

In case of identical sources, the contributions of all chains to cause an overflow are equally spread. The model can be reduced to a single source model in a straightforward fashion, see Ridder and Walrand [1992]. Here, we treat an example with two Markov modulated fluid model fed by two different sources.

We assume a chain on $E_1 := \{1,2,3\}$ and another on $E_2 := \{1,2\}$, equipped with rates $c := 4$, $\tilde{r}_1 := (0,2,2)$ and $\tilde{r}_2 := (4,0)$. In the manner stated above, these chains may be combined to a single chain on $\{(1,1),(2,1),(3,1),(1,2),(2,2),(3,2)\}$, with traffic rates $(0,2,2,-4,-2,-2)$. The transition matrices are given by

$$\Lambda^{(1)} := \begin{pmatrix} * & 1 & 0 \\ 0 & * & 2 \\ 4 & 0 & * \end{pmatrix}, \quad \Lambda^{(2)} := \begin{pmatrix} * & 2 \\ 1 & * \end{pmatrix}.$$

Λ is the transition matrix constructed by superposition of the independent chains $X_1(\cdot)$ and $X_2(\cdot)$. From the dual transition matrix M of Λ, the matrices $M^{(1)}$ and $M^{(2)}$ can be

extracted easily:

$$\Lambda = \begin{pmatrix} * & 1 & 0 & 2 & 0 & 0 \\ 0 & * & 2 & 0 & 2 & 0 \\ 4 & 0 & * & 0 & 0 & 2 \\ 1 & 0 & 0 & * & 1 & 0 \\ 0 & 1 & 0 & 0 & * & 2 \\ 0 & 0 & 1 & 4 & 0 & * \end{pmatrix}, \quad M = \begin{pmatrix} * & 2.554 & 0 & 0.446 & 0 & 0 \\ 0 & * & 1.033 & 0 & 0.446 & 0 \\ 3.033 & 0 & * & 0 & 0 & 0.446 \\ 4.489 & 0 & 0 & * & 2.554 & 0 \\ 0 & 4.489 & 0 & 0 & * & 1.033 \\ 0 & 0 & 4.489 & 3.033 & 0 & * \end{pmatrix}.$$

This results in $\theta^* = 1.261$. We see that $\langle \pi_M, \tilde{r} \rangle = 5.175$ consists of the components $\langle \pi_{M^{(1)}}, \tilde{r}_1 \rangle = 1.536$ and $\langle \pi_{M^{(2)}}, \tilde{r}_2 \rangle = 3.639$. In other words: the first chain contributes for about 30% to the input during an overflow and the second for the remaining 70%. In the steady state this ratio was about $39\% - 61\%$. We finish our numerical investigations with the corresponding simulation results.

Table 3: Simulation results Example 3.

B	6.0		9.0		15.0	
	Fast	Direct	Fast	Direct	Fast	LD
# cycles	10^3	10^6	10^3	10^6	10^3	
$\widehat{\alpha(B)}$	$4.48 \cdot 10^{-4}$	$4.39 \cdot 10^{-4}$	$1.09 \cdot 10^{-5}$	$1.40 \cdot 10^{-5}$	$5.23 \cdot 10^{-9}$	
RE	7.8%	8.7%	7.9%	43.8%	7.6%	
$-\widehat{\theta^*}$	1.29	1.29	1.27	1.24	1.27	1.261
$\widehat{T(B)}$	5.31	5.03	7.99	7.25	13.11	
RE	4.2%	4.0%	3.6%	8.0%	3.2%	
$\widehat{T(B)}/B$	0.89	0.84	0.89	0.82	0.87	0.851
Time	4:70	8:22:18	5.23	8:35:96	7:12	

We remark that the simulation results presented in the three examples above suggest that the probability of an overflow cycle is asymptotically exponential with decay parameter θ^* and an unknown amplitude (or prefactor) $\eta > 0$:

$$\alpha(B) \exp(\theta^* B) \to \eta, \text{ where } B \to \infty.$$

Furthermore, the 'fast simulation' procedure for $\alpha(B)$ can be extended to a method for estimating the long-run fraction of time that the buffer contents equals B, cf. Mandjes [1993].

9 Some conclusions

WE HAVE FOCUSED on Markov modulated input processes of a continuous buffer system endowed with finite capacity B. The overflow probability emerging in this model, $\alpha(B)$, can be approached on different probabilistic levels, with different advantages and

disadvantages. It is true that the level 1 method - evaluated in (1) - yields the decay rate of $\alpha(B)$ in quite a direct way, but it does not provide us insight in how the underlying Markov chain behaves when the fluid model reaches overflow. To meet this objective, we discussed alternative expressions for the exponent of $\alpha(B)$ using both level 2 and level 3 large deviations, cf. equations (5) and (7).

Our main study was to apply these expressions for variance reduction objectives when executing Monte Carlo simulations in order to estimate the overflow probability. With the aid of large deviations theory, we changed the underlying probability model such that the negative drift of the fluid system was replaced by a positive one along the optimal trajectory that causes overflows. The new estimators have variance properties obviously superior to the direct Monte Carlo estimators: the required simulation time is persistently orders of magnitude smaller. 'Fast simulation' permits us to execute various tests on buffer sizes and traffic characteristics to gain insight in the consequences for overflow probabilities and times.

References

[1] V. Anantharam [1988]. How large delays build up in a GI/G/1 queue. *Queueing Systems,* Vol. 5, p. 345-368.

[2] D. Anick, D. Mitra and M.M. Sondhi [1982]. Stochastic theory of data-handling System with Multiple Sources. *The Bell System Technical Journal,* Vol. 61, p. 1871-1894.

[3] J.A. Bucklew [1990]. *Large Deviation techniques in Decision, Simulation and Estimation.* Wiley, New York.

[4] M. Cottrell, J.-C. Fort and G. Malgouyres [1983]. Large Deviations and Rare Events in the Study of Stochastic Algorithms. *IEEE Transactions on Automatic Control,* Vol. 28, p. 907-920.

[5] C. Courcoubetis, G. Kesidis, A. Ridder, J. Walrand, and R. Weber [1991]. Admission Control and Routing in ATM Networks Using Inferences from Measured Buffered Occupancy. Memorandum UCB/ERL M91/37, Electronics Research Labaratory, College of Engineering, University of California, Berkeley.

[6] A. Dembo and O. Zeitouni [1993]. *Large Deviations Techniques and Applications.* Jones and Bartlett, Boston.

[7] G. De Veciana, C. Olivier, and J. Walrand [1993]. Large Deviations of Birth Death Markov Fluids. *Probability in the Engineering and Information Sciences,* Vol. 7, p. 237-255.

[8] M.D. Donsker and S.R.S. Varadhan [1975]. Asymptotic Evaluation of Certain Markov Process Expectations for Large Time, I. *Communications on Pure and Applied Mathematics* 28, p. 1-47.

[9] R. S. Ellis [1985]. *Entropy, Large Deviations, and Statistical Mechanics.* Springer Verlag, Berlin.

[10] P.W. Glynn and D.L. Iglehart [1989]. Importance Sampling for Stochastic Simulations. *Management Science,* Vol. 35, p. 1367-1392.

[11] G. Kesidis and J. Walrand [to appear]. Relative entropy between Markov transition rate matrices. *IEEE Transactions on Information Theory.*

[12] M. Mandjes [1993]. Finite-capacity GI/G/1 queueing systems with buffer overflows. Research Memorandum 1993-48a, Dept. of Econometrics, Free University Amsterdam.

[13] D. Mitra [1988]. Stochastic Theory of a Fluid Model of Producers and Consumers Coupled by a Buffer. *Advances of Applied Probability,* Vol. 20, p. 646-676.

[14] S. Parekh and J. Walrand [1989]. A Quick Simulation Method for Excessive Backlogs in Networks of Queues. *IEEE Transactions on Automatic Control,* Vol. 34, p. 54-66.

[15] A. Ridder [1991]. Markov Fluid Models and Large Deviations. Memorandum UCB/ERL M91/1, Electronics Research Laboratory, College of Engineering, University of California, Berkeley.

[16] A. Ridder and J. Walrand [1992]. Some Large Deviations Results in Markov Fluid Models. *Probability in the Engineering and Information Sciences,* Vol. 6, p. 543-560.

[17] A. Ridder [1993]. Fast Simulation of Markov Fluid Models. Research Memorandum 1993-21, Dept. of Econometrics, Free University Amsterdam.

[18] D. Siegmund [1976]. Importance Sampling in the Monte Carlo Study of Sequential Tests. *The Annals of Statistics,* Vol. 4, p. 673-684.

[19] T.E. Stern and A.I. Elwalid [1991]. Analysis of Separable Markov-Modulated Rate Models for Information-Handling Systems. *Advances in Applied Probability,* Vol. 23, p. 105-139.

[20] J. Walrand [1988]. *An Introduction to Queueing Networks.* Prentice-Hall, New Jersey.

[21] A. Weiss [1986]. A New Technique of Analyzing Large traffic Systems. *Advances in Applied Probability,* Vol.18, p. 506-532.

1992-1	R.J. Boucherie N.M. van Dijk	Local Balance in Queueing Networks with Positive and Negative Customers
1992-2	R. van Zijp H. Visser	Mathematical Formalization and the Analysis of Cantillon Effects
1992-3	H.L.M. Kox	Towards International Instruments for Sustainable Development
1992-4	M. Boogaard R.J. Veldwijk	Automatic Relational Database Restructuring
1992-5	J.M. de Graaff R.J. Veldwijk M. Boogaard	Why Views Do Not Provide Logical Data Independence
1992-6	R.J. Veldwijk M. Boogaard E.R.K. Spoor	Assessing the Software Crisis: Why Information Systems are Beyond Control
1992-7	R.L.M. Peeters	Identification on a Manifold of Systems
1992-8	M. Miyazawa H.C. Tijms	Comparison of Two Approximations for the Loss Probability in Finite-Buffer Queues
1992-9	H. Houba	Non-Cooperative Bargaining in Infinitely Repeated Games with Binding Contracts
1992-10	J.C. van Ours G. Ridder	Job Competition by Educational Level
1992-11	L. Broersma P.H. Franses	A model for quarterly unemployment in Canada
1992-12	A.A.M. Boons F.A. Roozen	Symptoms of Dysfunctional Cost Information Systems
1992-13	S.J. Fischer	A Control Perspective on Information Technology
1992-14	J.A. Vijlbrief	Equity and Efficiency in Unemployment Insurance
1992-15	C.P.M. Wilderom J.B. Miner A. Pastor	Organizational Typology: Superficial Foursome of Organization Science?
1992-16	J.C. van Ours G. Ridder	Vacancy Durations: Search or Selection?
1992-17	K. Dzhaparidze P. Spreij	Spectral Characterization of the Optional Quadratic Variation Process
1992-18	J.A. Vijlbrief	Unemployment Insurance in the Netherlands, Sweden, The United Kingdom and Germany