Faculteit der Economische Wetenschappen en Econometrie

021 Serie Research Memoranda

1993

Fast Simulation of Markov Fluid Models

Ad Ridder

Research Memorandum 1993-21
mei 1993

vrije Universiteit *amsterdam*

Fast Simulation of Markov Fluid Models

Ad Ridder

Free University of Amsterdam [1]

April 20, 1993

[1]Full address: Dept. of Economy and Econometrics, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands, phone: (+31) 20 548 7070, fax: (+31) 20 646 1440, email: aridder@sara.nl

Abstract

In this paper we study the problem of finding variance reduction for estimating probabilities of rare events in Markov Fluid Models via Monte Carlo simulation. We propose to apply Large Deviations Theory to the processes for obtaining asymptotic expressions of these probabilities. Then we shall consider variance reduction by means of importance sampling where the new statistical law of the process is derived from the large deviations expressions.

MARKOV FLUID MODELS · OVERFLOW PROBABILITIES · LARGE DEVIATIONS · MONTE CARLO SIMULATION · IMPORTANCE SAMPLING

1 Introduction

Let $\{X_t\}_{t\geq 0}$ be a continuous time Markov chain on a finite state space $E = \{1, 2, \ldots, d\}$ with transition rates $\{q_{ij}\}_{i,j\in E}$, and let f be a real valued function on E. We assume that $\max_{i\in E} f(i) > 0$, that $\{X_t\}$ has stationary distribution π, and that $\sum_{i\in E} \pi_i f(i) < 0$. Fix $B > 0$ and define (for $t \geq 0$)

$$J_t = \min\left\{B, \max\left(0, \int_0^t f(X_s)ds\right)\right\} \tag{1}$$

The process $\{J_t\}_{t\geq 0}$ describes the contents of a finite buffer in a so-called Markov Fluid Model. That is an input-output system with an incoming stream of packets at rate r_i when the chain $\{X_t\}$ is in state i. The packets flow in a finite buffer of size B that is emptied at a constant rate of c packets. When the input rate exceeds the output rate, the buffer is filled eventually upto its capacity B. Packets that arrive while the buffer is full, are lost. When we define the flow function f by $f(i) = r_i - c$, we obtain the process $\{J_t\}$ of (1) that describes the number of packets in the buffer at any time t. Equivalently we may write

$$J_{t+dt} = J_t + f(X_t)dt$$

whenever the right-hand side is feasible (i.e. $0 \leq J_t + f(X_t)dt \leq B$), otherwise $J_{t+dt} = J_t$.

Markov Fluid Models are proposed for modeling buffer behaviours in switches of high speed communication networks [2, 4, 11, 15]. The idea is that traffic comes mainly in bursts or peaks alternating with quiet moments. 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 rate function r for the generation of packets that the customer wishes to send through the network. When several customers or connections share the same buffer, we may superimpose their chains and rate functions into one chain and rate function.

Given the traffic characteristics in terms of a chain $\{X_t\}$ and the rate functions r and c, we shall consider the estimation of the probability that a buffer overflow will occur. We characterise this probability in the following way. When the buffer process evolves in time it generates cycles. A cycle consists of a busy period and an idle period (similarly to e.g. $M/G/1$). However notice that during an idle period packets may arrive at the buffer, viz. at a rate less than c. We discern two different cycles to indicate whether an overflow occurred during the busy period or not: overflow and regular cycles. Let $E^+ \subset E$ contain all states i for which $f(i) > 0$. 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 α_i. Now let us define the quasi-stationary distribution

$$\nu_i = \frac{\pi_i}{\sum_{j \in E^+} \pi_j}$$

for $i \in E^+$, then our overflow probability is

$$\alpha_B = \sum_{i \in E^+} \nu_i \alpha_i \tag{2}$$

A second issue that attracts attention in these models, is the expected time between two overflow events [1, 12, 17]. If we assume that the buffer contents has returned to 0 in between, we have to wait for an overflow cycle. In the overflow cycle the buffer builds up from 0 to B. We call the time during which this happens, the overflow time. We are interested in estimating its expectation τ_B. Notice that similarly to the evaluation of the overflow probability the expected overflow time may be defined by

$$\tau_B = \sum_{i \in E^+} \nu_i \tau_i \tag{3}$$

where τ_i is the expected overflow time in an overflow cycle that starts with the chain $\{X_t\}$ in state i.

We have organized this paper as follows. Section 2 presents Large Deviations expressions of the overflow probability and the overflow time. In Section 3 we use

these expressions for getting variance reductions in Monte Carlo simulations. Our approach is to apply Importance Sampling and we need the Large Deviations results for deriving an optimal change of measure within the class of exponentially twisted probabilities. Section 4 contains two numerical examples. In Section 5 we attack the problem of multi-input of which the basics are actually done in Sections 2 and 3.

2 Large Deviations expressions

In [13] we have shown the following asymptotic expression for the overflow probability

$$\lim_{B\to\infty} \frac{1}{B} \log \alpha_B = -\theta^* \tag{4}$$

where θ^* can be found in two ways, based on the so-called level 1 and level 2 Large Deviation Principle [7]. For that purpose, consider the chain $\{X_t\}$ and its return times $\{Y_n\}_{n\geq 0}$ to some fixed chosen state i^*. Assume that the buffer has no boundaries at 0 and B, so we define the free process

$$\bar{J}_t = \int_0^t f(X_s)ds$$

for decribing the contents of a virtual (or free) buffer. The free buffer increments between consecutive return times are i.i.d. random variables:

$$\xi_n = \bar{J}_{Y_n} - \bar{J}_{Y_{n-1}} \tag{5}$$

Notice that the actual buffer and the virtual buffer behave similarly during the overflow time in an overflow cycle. Let $I^{(1)}(\cdot)$ be the Legendre-Fenchel Transform (LFT) of the cumulant $\log \mathsf{E} \exp(\theta\xi_1)$:

$$I^{(1)}(x) = \sup_{\theta\in\mathbf{R}} \{\theta x - \log \mathsf{E} \exp(\theta\xi_1)\} \tag{6}$$

Then

$$\theta^* = \theta^{(1)} = \inf\ \tau I^{(1)}\left(\frac{1}{\tau}\right) \tag{7}$$

where infimum is taken from all $\tau > 0$. The derivation of this expression is based on the Large Deviations Principle [5, 9, 18]

$$\theta^* = \inf \int_0^{T(\phi)} I^{(1)}(\phi'(t))dt$$

where the infimum is taken from nonnegative continuous differentiable (almost everywhere) functions ϕ with $\phi(0) = 0$ and $T(\phi) = \inf\{t > 0 : \phi(t) = 1\}$. We find (7) by applying convex analysis arguments. It says that the optimal function or 'path' ϕ is a straight line with positive slope which has to be determined in (7).

When we recall the actual Fluid Model, we can heuristically argue that during the overflow time the chain $\{X_t\}$ typically behaves according to some distribution μ^* rather than to the stationary π. Write η^* for the expected return times to the chosen state i^* under μ^*:

$$\eta^* = \mathsf{E}_{\mu^*}(Y_n - Y_{n-1})$$

and let $\tau^{(1)}$ be the minimizer of the right-hand side of (7). Using properties of the LFT one can show that there is a unique minimum. Then $\tau^{(1)}$ equals (approximately) the expected number of returns to state i^* before the process reaches level 1. So (for large B)

$$\tau_B \approx B\eta^*\tau^{(1)}$$

Furthermore, the buffer process $\{J_t\}$ follows during the overflow time approximately a straight line with slope

$$\left(\eta^*\tau^{(1)}\right)^{-1}$$

A second expression of θ^* uses Large Deviations for empirical distributions. Define for any probability measure μ on E the entropy function [6, 16]

$$I^{(2)}(\mu) = -\inf_{u>0} \sum_{i,j\in E} \mu_i q_{ij} \frac{u_j}{u_i} \tag{8}$$

(where $u = (u_1, \ldots, u_d) > 0$ componentwise), and denote as usual the inner product

$$\langle \mu, f \rangle = \sum_{i \in E} \mu_i f(i) = \sum_{i \in E} \mu_i r_i - c$$

Then again from the Large Deviations Principle and from convex analysis [13]

$$\theta^* = \theta^{(2)} = \inf \ \tau I^{(2)}(\mu) \tag{9}$$

where infimum is taken from all $\tau > 0$ and probability measures μ on E such that

$$\langle \mu, f \rangle = \frac{1}{\tau} \tag{10}$$

Here the interpretation goes along the following lines. Let $\tau^{(2)}$ and $\mu^{(2)}$ be the arguments that mimimize the right-hand side of (9) – these exist –. Then $\mu^{(2)}$ is the most likely distribution according to which the chain $\{X_t\}$ behaves during overflow time, in the previous paragraph called μ^*. The buffer will be filled with a speed of $\langle \mu^{(2)}, f \rangle$ per unit of time, hence from (10) $\tau^{(2)}$ is the expected time until level 1 is reached. So for large B

$$\tau_B \approx B\tau^{(2)}$$

The path that the buffer process will follow most likely during overflow time is a line with slope

$$\left(\tau^{(2)}\right)^{-1}$$

In [13] (Theorem 3) we have shown formally that indeed

$$\begin{aligned} \theta^{(2)} &= \theta^{(1)} \\ \tau^{(2)} &= \tau^{(1)}\eta^{(2)} \end{aligned} \tag{11}$$

where $\eta^{(2)}$ is the expected return time under $\mu^{(2)}$, previously denoted by η^*. In the following section we shall prove that the distribution $\mu^{(2)}$ is optimal with respect to a variance reduction objective.

3 Simulation and Importance Sampling

Now suppose that we wish to estimate the overflow probability α_B by using Monte Carlo simulation. For that purpose assume some underlying probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and

$$\alpha_B = \mathsf{P}(A)$$

with A the rare event of an overflow. 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 B. In the last case ω belongs to A. An arbritray ω is of the form

$$\omega = ((i_0, t_0), (i_1, t_1), \ldots, (i_n, t_n)) \tag{12}$$

Each i_k indicates the state of the chain after the k-th jump during the sample cycle and t_k measures the length of staying in i_k. Clearly the cycle starts in a state $i_0 \in E^+$, and since $\sum_{k=0}^{r} f(i_k)t_k$ stands for the buffer contents just before the $r+1$-th jump, $0 < \sum_{k=0}^{r} f(i_k)t_k < B$ holds for $r = 0, 1, \ldots, n-1$. And then either $\sum_{k=0}^{n} f(i_k)t_k \leq 0$ in which case $\omega \notin A$, or $\sum_{k=0}^{n} f(i_k)t_k \geq B$ in which case $\omega \in A$. It is possible that $n = \infty$ but almost all ω have finitely many jumps because the buffer process has a negative drift.

Suppose that $\omega \in A$ is of the form (12). The probability density of ω is

$$d\mathsf{P}(\omega) = \left(q_{i_n} \prod_{k=0}^{n-1} q_{i_k i_{k+1}} \right) \exp\left(- \sum_{k=0}^{n} q_{i_k} t_k \right) \; dt_0 \cdots dt_n \tag{13}$$

(with as usual $q_i = \sum_{j \neq i} q_{ij}$). Notice that $i_n \in E^+$ to ensure that an overflow can occur while the chain stays in that state.

For Monte Carlo simulations we draw arbritary ω's of the form (12) and estimate α_B by the relative frequency of the number of occurences in A. For large levels B the

number of samples to draw must be large in order to obtain good relative efficiency of the estimate since it is of the order of $1/\alpha_B$. E.g. when the confidence should be 95% and the efficiency – i.e. relative width of confidence interval – should be 10%, this number is approximately $400/\alpha_B$. Consider then the possibility of doing the simulations based on another probability Q, such that P is absolute continuous with respect to Q, a technique called Importance Sampling (e.g. [10, 14]). Let L be a Radon Nikodym derivative or likelihood function:

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

If we draw m samples ω respectively in the original (with P) and in the new simulations (with Q) the estimator has variances

$$\frac{1}{m}\left(\int_A d\mathsf{P} - \alpha_B^2\right) \text{ and } \frac{1}{m}\left(\int_A L^2 d\mathsf{Q} - \alpha_B^2\right) \tag{14}$$

Since $\int_A L^2 d\mathsf{Q} = \int_A L d\mathsf{P}$ we get immediately a necessary condition of variance reduction in terms of the likelihood L

$$\int_A L d\mathsf{P} < \int_A d\mathsf{P}$$

Secondly we may try to derive from (14) an optimal Q, i.e. the one that minimizes $\int_A L^2 d\mathsf{Q}$. It turns out that the optimal Q is not practical [3] and therefore we seek an optimal Q in the class of exponentially twisted (or tilted) probabilities [14]. This class contains probabilities parameterized by $\theta \in \mathbb{R}$, so we write P^θ for a typical member. Under P^θ the chain $\{X_t\}$ has transition rates $\{q_{ij}^\theta\}$ and the the free buffer increments $\{\xi_n\}$ (see (5)) have cumulative distribution function F^θ. Under the original probability P the increments have distribution F. Then we characterise our class of twisted probabilities by requiring that P^θ belongs to this class if and only if

$$dF^\theta(x) = \frac{\exp(\theta x) dF(x)}{\int \exp(\theta x) dF(x)} \text{ and } \int_{\mathbb{R}} \exp(\theta x) dF(x) < \infty \tag{15}$$

Let $\Theta \subset \mathbb{R}$ be the parameter space of all those θ for which P^θ satisfies (15). Notice that $0 \in \Theta$, i.e. the original probability P is an element of the class.

Now Theorem 2 of [5] and Theorem 1 of [13] provide us with the optimal P^θ in the sense of minimizing

$$\int_A L^2 d\mathsf{P}^\theta \quad (L = \frac{d\mathsf{P}}{d\mathsf{P}^\theta},\ \theta \in \Theta)$$

The optimal twisted probability is $\mathsf{P}^* = \mathsf{P}^{\theta^*}$ with θ^* satisfying (7) (or (9)). When we actually execute the optimization in (6) and (7) we find that

$$\int_{\mathsf{R}} \exp(\theta^* x) dF(x) = 1 \tag{16}$$

So clearly $\theta^* \in \Theta$. The left hand side of (16) is the moment generating function of the increments ξ_n evaluated in θ^*. In some "simple" models the moment generating function can be determined in closed expression. Then equation (16) may be appropriate for calculating the optimal θ^* (see the examples in Section 4).

At this stage we are left with the task of finding transition rates $\{q^*_{ij}\}$ that go with the optimal P^*. After we have implemented these rates we may execute the "quick" simulations and obtain the best variance reduction within the class of exponentially twisted. First we shall show that the "optimal" stationary distribution π^* of the chain under P^* is the "most likely" distribution $\mu^{(2)}$ that we found from the optimization program (9).

Lemma 1 $\quad \pi^* = \mu^{(2)}$

Proof. Denote the Large Deviation rate function $I^{(1)}$, expressed in (6), by $I^{(1)}_\xi$ and the rate function $I^{(2)}$, expressed in (8), by $I^{(2)}_X$. The subscripts ξ and X indicate the fact that the functions concern respectively the process $\{\xi_n\}$ and the chain $\{X_t\}$. The superscripts (1) and (2) indicate Large Deviations Principles for respectively sample means and empirical distributions.

Also it is possible to derive a Large Deviations Principle for empirical distributions, or equivalently for their associated distribution functions, of the process $\{\xi_n\}$ (e.g.

[7] ch VIII). Denote the rate function by $I_\xi^{(2)}$. According to the contraction principle the relation

$$I_\xi^{(1)}(x) = \inf\{I_\xi^{(2)}(G) : G \text{ distribution function on } \mathbf{R} \text{ with } \int_{\mathbf{R}} u dG(u) = x\}$$

holds. The infimum is attained at a distribution F^θ of the form (15) for a unique θ. When we execute the optimizations of (6) and (7) and apply the contraction principle for $x = 1/\tau^{(1)}$, we find the optimal distribution to be precisely $F^* = F^{\theta^*}$ with θ^* as in (7) (or (9)).

Suppose that the chain $\{X_t\}$ has stationary distribution μ, assuming a probability on the sample space which is not further specified. By η we denote the expected average return times to a fixed state i^* and by G the distribution function of the free buffer increments $\{\xi_n\}$. Then (see Lemma 5 in [13])

$$\begin{aligned} \int_{\mathbf{R}} u dG(u) &= \eta \langle \mu, f \rangle \\ I_\xi^{(2)}(G) &= I_X^{(2)}(\mu)\eta \end{aligned}$$

We particularly concentrate on the triplets $(\mu^{(2)}, \eta^{(2)}, G^{(2)})$ and (π^*, η^*, F^*). Here $\mu^{(2)}$ is extracted from (9) to be the minimizer of

$$\frac{I_X^{(2)}(\mu)}{\langle \mu, f \rangle}$$

Hence

$$\frac{I_X^{(2)}(\pi^*)}{\langle \pi^*, f \rangle} \geq \frac{I_X^{(2)}(\mu^{(2)})}{\langle \mu^{(2)}, f \rangle}$$

Therefore

$$\begin{aligned} I_\xi^{(2)}(F^*) &= I_X^{(2)}(\pi^*)\eta^* \\ &= \frac{I_X^{(2)}(\pi^*)}{\langle \pi^*, f \rangle} \int_{\mathbf{R}} u dF^*(u) \\ &\geq \frac{I_X^{(2)}(\mu^{(2)})}{\langle \mu^{(2)}, f \rangle} \int_{\mathbf{R}} u dF^*(u) \\ &= I_\xi^{(2)}(G^{(2)}) \frac{\int u dF^*(u)}{\eta^{(2)} \langle \mu^{(2)}, f \rangle} \end{aligned}$$

By definition, $\int u dF^* = 1/\tau^{(1)}$, and from (9) and (11)

$$\eta^{(2)}(\mu^{(2)}, f) = \frac{\eta^{(2)}}{\tau^{(2)}} = \frac{1}{\tau^{(1)}}$$

So

$$I_\xi^{(2)}(F^*) \geq I_\xi^{(2)}(G^{(2)})$$

The $\leq$-inequality follows immediately from

$$\int_{\mathbb{R}} u dG^{(2)}(u) = \eta^{(2)}(\mu^{(2)}, f) = \frac{1}{\tau^{(1)}}$$

and from the definition of F^*. Because F^* is the unique optimum of the contraction principle, we must have that $G^{(2)} = F^*$ and therefore also $\mu^{(2)} = \pi^*$. □

Finally we shall present sufficient conditions for the transition rates $\{q_{ij}^*\}$ to fulfill in order of getting the optimal change of measure $\mathsf{P} \rightarrow \mathsf{P}^*$. Assume some probability $\tilde{\mathsf{P}}$. It induces transition rates $\{\tilde{q}_{ij}\}$ of the chain $\{X_t\}$ and a distribution function $\tilde{F}$ of the increments $\{\xi_n\}$. Recall that the original P is given by way of known rates $\{q_{ij}\}$ (see (13)).

Lemma 2 *If $\{\tilde{q}_{ij}\}$ satisfy*
(i) for any $i \in E$

$$\tilde{q}_i = q_i - \theta^* f(i)$$

(ii) for any feasible cycle of states $i_0 = i^*, i_1, \ldots, i_r, i_{r+1} = i^*$, meaning $i_1, \ldots, i_r \neq i_0$ and $q_{i_k i_{k+1}} > 0$ for $k = 0, 1, \ldots, r$,

$$\prod_{k=0}^{r} \tilde{q}_{i_k i_{k+1}} = \prod_{k=0}^{r} q_{i_k i_{k+1}}$$

Then

$$\tilde{F} = F^*$$

Proof. Let $i_0 = i^*, i_1, \ldots, i_r, i_{r+1} = i^*$ be a feasible cycle and assume that the chain stays a time t_k in state i_k. The probability density of this realisation to occur is

$$\left(\prod_{k=0}^{r} q_{i_k i_{k+1}}\right) \exp\left(-\sum_{k=0}^{r} q_{i_k} t_k\right) \quad dt_0 dt_1 \cdots dt_r$$

Using (i) and (ii) of Lemma 2 this density equals

$$\left(\prod_{k=0}^{r} \tilde{q}_{i_k i_{k+1}}\right) \exp\left(-\sum_{k=0}^{r} \tilde{q}_{i_k} t_k - \theta^* \sum_{k=0}^{r} f(i_k) t_k\right) \quad dt_0 dt_1 \cdots dt_r$$

The event $\xi_n \in (x, x + dx)$ is made up of all feasible cycles of the form given and of all duration times t_k such that $\sum_{k=0}^{r} f(i_k) t_k = x$. Summing all the corresponding densities, assuming respectively P and $\tilde{\mathsf{P}}$,

$$dF(x) = \exp(-\theta^* x) d\tilde{F}(x)$$

or, using (16),

$$d\tilde{F}(x) = \exp(\theta^* x) dF(x) = dF^*(x)$$

□

After we have determined transition rates $\{q_{ij}^*\}$ such that (i) and (ii) of Lemma 2 are fulfilled, we may run simulations by drawing sample ω's using the probability P^* and compensate their occurences by the likelihood $L^* = \frac{d\mathsf{P}}{d\mathsf{P}^*}$. Recall that when $\omega \in A$ it induces an overflow. Implementing $d\mathsf{P}(\omega)$ as in (13), a similar expression of $d\mathsf{P}^*(\omega)$, and applying (i) of Lemma 2, we get for $\omega \in A$

$$L^*(\omega) = \left(\frac{q_{i_n} \prod_{k=0}^{n-1} q_{i_k i_{k+1}}}{q_{i_n}^* \prod_{k=0}^{n-1} q_{i_k i_{k+1}}^*}\right) \exp(-\theta^* B)$$

The drift of a fluid process is the average net amount of fluid per unit of time, originally $\sum_{i \in E} \pi_i f(i) < 0$. Under P^* the drift becomes (notation as in Lemma 1)

$$\sum_{i \in E} \pi_i^* f(i) = \frac{\int u dF^*}{\eta^*} = \frac{1}{\tau^{(2)}} > 0$$

This is the slope of the "most likely" path causing overflows.

4 Examples

In this section we work out some of the concepts of the previous sections.

Example A

The most simple fluid model consists of a chain with two states, $E = \{1, 2\}$. So $q_1 = q_{12}, q_2 = q_{21}$. Let $f(1) < 0 < f(2)$ meaning that state 1 represents the quiet moments and state 2 the bursty moments during a communication connection. Notice that

$$\pi = \frac{1}{q_1 + q_2}(q_2, q_1) \text{ and } \langle \pi, f \rangle < 0$$

gives $q_1 f(2) + q_2 f(1) < 0$. After some algebra we find

$$\begin{aligned}
\theta^* &= \frac{q_1}{f(1)} + \frac{q_2}{f(2)} \\
\tau^{(1)} &= -\frac{q_1 q_2}{q_1 f(2) + q_2 f(1)} \\
q_1^* &= q_2 \frac{-f(1)}{f(2)} \\
q_2^* &= q_1 \frac{f(2)}{-f(1)}
\end{aligned}$$

Then based on Lemmata 1 and 2

$$\begin{aligned}
\mu^{(2)} &= \pi^* = \frac{1}{q_1^* + q_2^*}(q_2^*, q_1^*) \\
\tau^{(2)} &= (\langle \pi^*, f \rangle)^{-1} = \frac{q_1 \frac{f(2)}{f(1)} + q_2 \frac{f(1)}{f(2)}}{q_1 f(2) + q_2 f(1)} \\
L^*(\omega) &= \frac{q_2}{q_2^*} \exp(-\theta^* B), \ \omega \in A
\end{aligned}$$

Table 1 contains the results of simulations that were run on the model with $q_1 = 10, q_2 = 30, f(1) = -1100, f(2) = 2500$. The drift of the system is originally -200 and after changing the rates to $q_1^* = 13.2, q_2^* = 22.73$ it is 222.57. We run a number of cycles (NC) until the 95% confidence interval of the estimated overflow probability

$\hat{\alpha}_B$ has relative width of near 10% to each side, called the relative efficiency (RE). In stead of the estimated overflow time $\hat{\tau}_B$ we present in Table 1 the normalized overflow time $\hat{\tau}_B/B$ (also with relative efficiency of 95% confidence). Furthermore we tabulate the fraction of time of staying in state 1 during overflow time, which estimates the empirical distribution $\mu^{(2)}$ (state 2 is omitted since its probability is simply the complement). No efficiencies are given there but these are smaller than those of the overflow times. Each buffer size is run twice, once with "direct" simulations and once with "quick" simulations. The last column of the table contains the values based on the Large Deviations expressions given above.

B	2000		2500		3000		LD
NC	900K	1500	3M	1500	11M	1500	
$\hat{\alpha}_B \times 10^4$	6.400	9.131	1.447	2.224	0.301	0.501	
RE	8.16	9.19	9.41	8.94	10.77	9.16	
$(-\log \hat{\alpha}_B/B) \times 10^3$	3.677	3.499	3.536	3.364	3.370	3.301	2.909
$(\hat{\tau}_B/B) \times 10^3$	3.294	3.089	3.524	3.549	3.889	3.497	4.491
RE	4.34	6.15	4.77	5.73	4.93	5.16	
$\hat{\mu}_1^{(2)}$	0.610	0.605	0.616	0.616	0.623	0.615	0.633

Table 1: Direct and quick simulation estimates in Example A. NC means number of cycles, K thousand, M million, RE relative efficiency (of the 95% confidence interval) in % of the estimate given just above it.

A couple of remarks: in the quick simulations case the number NC does not grow with B because the likelihood ratio L takes care of that. The quick estimates of the overflow probability are persistently larger than the direct ones and even the confidence intervals do not overlap. The relative difference in the three cases is 42%, 53% and 66%. Unclear why this phenomenon happens here and not in the following examples. The estimates of overflow times and empirical distribution do match quite

well: relative differences of at most 10% but in most cases much smaller. The Large Deviations expressions of (i) the normalized logarithm of the overflow probability, (ii) the overflow time and (iii) the empirical distribution are "quite different" from the estimates given here. Relative differences are in case (i) between 13 and 26%, in case (ii) between 13 and 31%, and in case (iii) between 1.5 and 4.5%. The Large Deviations expressions are asymptotic results when $B \to \infty$ and we expect that estimates for larger buffer sizes should become closer to these. This is done in Table 2.

B	$(-\log \hat{\alpha}_B/B) \times 10^3$	$(\hat{\tau}_B/B) \times 10^3$	$\hat{\mu}_1^{(2)}$
5000	3.138	3.818	0.621
10000	3.028	4.131	0.627
15000	2.988	4.318	0.630
20000	2.964	4.407	0.631
25000	2.954	4.380	0.631
30000	2.950	4.392	0.631
LD	2.909	4.491	0.633

Table 2: Quick simulation estimates for large buffer sizes in Example A.

Example B In the second example the chain has three states with $f(1) < f(2) < f(3)$: state 1 represents light loaded, state 2 moderate loaded and state 3 heavy loaded traffic. We assume that the chain only jumps between states 1 and 2, and between states 2 and 3: $q_1 = q_{12}, q_{13} = 0, q_3 = q_{32}, q_{31} = 0$. We can use equation (16) to solve for θ^*. After some algebra we find that θ^* is the (unique) positive root of

$$a_1 a_2 a_3 \theta^2 - (a_1 a_2 + a_1 a_3 + a_2 a_3)\theta + (a_2 + a_3 + a_1 p - a_3 p) = 0$$

where

$$a_i = \frac{f(i)}{q_i} \text{ and } p = \frac{q_{21}}{q_2}$$

The optimal rates $\{q^*_{ij}\}$ follow from Lemma 2: $q^*_{12} = q^*_1 = q_1 - \theta^* f(1)$ and $q^*_{12}q^*_{21} = q_{12}q_{21}$. Similarly for q^*_{32}, q^*_{23}. From these the stationary distribution $\pi^* = \mu^{(2)}$ is determined and the drift $\langle \pi^*, f \rangle$. Suppose $f(1) < 0 < f(2) < f(3) : E^+ = \{2,3\}$, then cycles of the buffer process may start in state 2 or 3, overflow can occur while the chain stays in 2 or 3. The likelihood ratio $L^*(\omega)$ on A takes on the form $H(\omega)\exp(-\theta^* B)$ with $H(\omega)$ different in these four cases.

Table 3 contains results of simulations that were run on the model with $q_1 = 10, q_{21} = 20, q_{23} = 30, q_3 = 40, f(1) = -1500, f(2) = 500, f(3) = 1500$. The drift of the system is originally -366.67 and after changing the rates to $q^*_1 = 16.33, q^*_{21} = 12.24, q^*_{23} = 35.64, q^*_3 = 33.66$ it is 343.15.

B	1500		1750		2000		LD
NC	500K	500	1.5M	500	3.5M	500	
$\hat{\alpha}_B \times 10^4$	7.700	9.883	2.567	2.781	0.966	1.007	
RE	9.99	9.03	9.99	10.91	10.66	10.51	
$(-\log \hat{\alpha}_B)/B \times 10^3$	4.779	4.613	4.724	4.679	4.623	4.602	4.223
$\hat{\tau}_B/B \times 10^3$	2.328	2.210	2.436	2.442	2.424	2.516	2.914
RE	4.17	5.11	4.14	6.23	4.33	6.06	
$\hat{\mu}^{(2)}_1$	0.234	0.226	0.240	0.242	0.240	0.248	0.267
$\hat{\mu}^{(2)}_2$	0.369	0.370	0.369	0.364	0.368	0.359	0.356
$\hat{\mu}^{(2)}_3$	0.397	0.404	0.391	0.394	0.392	0.393	0.377

Table 3: Direct and quick simulation estimates in Example B.

The same remarks as above in Example A can be made here, except that the results

show "better": the relative differences between the direct and quick estimates, and between the estimates and the asymptotic expressions are less here, except for the empirical distribution estimates. This cannot be explained by saying that the buffer sizes in Example B lie "closer to infinity" because we observe that the overflow probabilities of the two examples are of the same order.

5 Multi input

In Section 1 we associated a customer who is connected to a communication network and who loads packets into a buffer of finite capacity, with a Markov chain $\{X_t\}$ and a rate function r on the state space of $\{X_t\}$. The chain describes the time behaviour of the connection and the rate function reflects the loading characteristics. The buffer is emptied at a constant rate c. In this section 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 $\{X_t(k)\}$ on a (finite) state space $E(k)$ with transition rates $\{q_{ij}(k)\}$, and by an input rate function $r(k)$ on $E(k)$. It is a matter of a simple transformation to obtain the model of Section 1. The (vector) process $X_t = (X_t(1), \ldots, X_t(K))$ is a Markov chain on $E = E(1) \times \cdots \times E(K)$. Transitions take place only when one of the components changes (the event of two or more simultaneous changes has probability 0). The flow function f on E becomes (for $i = (i(1), \ldots, i(K))$) $f(i) = \sum_{k=1}^{K} r_{i(k)}(k) - c$.

Again we are interested in estimating the overflow probability α_B and the expected overflow time τ_B via Monte Carlo simulation using importance sampling. We write $I^{(2)}(k)(\mu(k))$ for the entropy function (8) applied to the probability measure $\mu(k)$ on $E(k)$ and using rates $\{q_{ij}(k)\}$. The product measure $\mu = (\mu(1), \cdots, \mu(K))$ on E may be interpreted as an empirical distribution of the chain X_t. Then by a straightforward extension of Lemma 6 of [13] we obtain again (4) with

$$\theta^* = \inf \tau \sum_{k=1}^{K} I^{(2)}(k)(\mu(k)) \tag{17}$$

where infimum is taken from all $\tau > 0$ and probability measures $\mu = (\mu(1), \cdots, \mu(K))$ on E such that

$$\sum_{k=1}^{K} \langle \mu(k), r(k) \rangle - c = \frac{1}{\tau} \tag{18}$$

Solving this optimisation program we find again optimal $\tau^{(2)}$ and $\mu^{(2)}$ such that the expected overflow time τ_B for large B is approximately $B\tau^{(2)}$ and the k-th chain behaves according to the marginal $\mu^{(2)}(k)$. A remarkable reduction property is present in models with identical inputs, i.e. all chains have the same transition rates – say $\{q_{ij}\}$ – and all input rate functions are identical – say r_i –, and says that the contribution of all chains to cause an overflow is equally spread [4, 13],

$$\langle \mu^{(2)}(k), r \rangle - \frac{c}{K} = \frac{1}{\tau^{(2)}} \tag{19}$$

and that all optimal marginals $\mu^{(2)}(k)$ are the same as well. That means that the estimation of the overflow probability and of the overflow time using Monte Carlo simulations may be executed in a single input model by replacing the output rate c by c/K.

In case that not all input sources are identical, the analysis of Section 3 still holds and may be executed here to obtain variance reduction in simulations. In particular Lemmata 1 and 2 are applicable in the multi-dimensional setting. Omitting (numerical) details we present below the results for a specific model.

Example C

Consider the fluid model with two (independent) sources both consisting of two states (numbered 1 and 2) and with the same input rate function r. In the simulations we take $q_1(1) = 10, q_2(1) = 30$, $q_1(2) = 1, q_2(2) = 3$, $r_1 = 100, r_2 = 2100$ and $c = 1500$. Notice that the steady state distribution of the two chains are identical ($\pi(k) = (0.75, 0.25)$) and that both contribute an input of 600 packets per time unit. The equilibrium drift becomes therefore -300 packets per time unit. After

solving numerically (17) and (i) and (ii) of Lemma 2 we obtain the optimal twisted rates $q_1^*(1) = 10.61, q_2^*(1) = 28.27$ $q_1^*(2) = 1.91, q_2^*(2) = 1.57$ and drift 342.08 packets per time unit. Table 4 shows the simulation results and the corresponding values according to Large Deviations. We observe that overflows are mainly caused by "non-equilibrium" behaviour of the second chain.

B	5000		6000		7500		LD
NC	600K	750	1.5M	750	9M	750	
$\hat{\alpha}_B \times 10^4$	7.517	18.572	2.433	5.508	0.422	0.932	
RE	9.23	8.97	10.26	9.34	10.05	9.43	
$(-\log \hat{\alpha}_B/B) \times 10^3$	1.439	1.258	1.387	1.251	1.343	1.237	1.168
$(\hat{\tau}_B/B) \times 10^3$	2.188	2.133	2.329	2.277	2.362	2.425	2.923
RE	4.73	5.28	5.14	5.36	4.80	5.59	
$\hat{\mu}_1^{(2)}(1)$	0.722	0.724	0.725	0.723	0.723	0.725	0.727
$\hat{\mu}_1^{(2)}(2)$	0.399	0.392	0.411	0.408	0.415	0.419	0.452

Table 4: Direct and quick simulation estimates in Example C.

Again the quick simulations lead to larger estimates of the overflow probabilities (in fact twice as large). We expect better performances for larger buffer sizes. Table 5 shows the convergence of the estimates to the Large Deviations values.

6 Conclusions

We have focused on Markov modulated input processes of a continuous buffer system. The overflow probability (2) satisfies the asymptotic expression (4) for various models. Based on Large Deviations Principles the asymptotic expression may be

B	$(-\log \hat{\alpha}_B/B) \times 10^3$	$(\hat{\tau}_B/B) \times 10^3$	$\hat{\mu}_1^{(2)}(1)$	$\hat{\mu}_1^{(2)}(2)$
10000	1.215	2.564	0.724	0.431
20000	1.193	2.767	0.727	0.442
30000	1.186	2.818	0.724	0.448
40000	1.179	2.780	0.727	0.443
50000	1.177	2.903	0.727	0.451
60000	1.176	2.805	0.727	0.445
LD	1.168	2.923	0.727	0.452

Table 5: Quick simulation estimates for large buffer sizes in Example C.

evaluated by (7),(9) or (16). Our main study was to apply this expression for variance reduction purposes when executing Monte Carlo simulations in order to estimate the overflow probability. With the aid of the Large Deviations expression we can change the probabilty measure so that the negative drift of the system (in equilibrium) becomes a positive one along the optimal path that causes overflows. The quick simulation estimates make it possible 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. On Fast Simulation of the Time to Saturation of Slotted ALOHA. *Journal of Applied Probability*, Vol. 29, p. 682 – 690, 1092.

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

[3] J.A. Bucklew, P. Ney and J.S. Sadowsky. Monte Carlo Simulation and Large Deviations Theory for Uniformly Recurrent Markov Chains. *Journal of Applied Probability*, Vol. 27, p. 44 – 59, 1990.

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

[5] M. Cottrell, J-C. Fort and G. Malgouyres. Large Deviations and Rare Events in the Study of Stochastic Algorithms. *IEEE Transactions on Automatic Control*, Vol. 9, p. 907 – 920, 1983.

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

[7] R.S. Ellis. *Entropy, Large Deviations and Statistical Mechanics.* Springer, New York, 1985.

[8] R.S. Ellis. Large Deviations for a General Class of Random Vectors. *Annals of Probability*, Vol. 12, p. 1 – 12, 1988.

[9] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems.* Springer, New York, 1984

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

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

[12] S. Parekh and J. Walrand. A Quick Simulation Method for Excessive Backlog in Network of Queues. *IEEE Transactions on Automatic Control*, Vol. 34, p. 54 – 66, 1989.

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

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

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

[16] S.R.S. Varadhan. *Large Deviations and Applications. CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, Pa, 1984.

[17] A. Weiss. A New Technique of Analyzing Large Traffic Systems. *Advances of Applied Probability*, Vol. 18, p.506 – 532, 1986.

[18] A.D. Wentzell. Rough Limit Theorems on Large Deviations for Markov Stochastic Processes, part II. *Theory of Probability and its Applications*, Vol 21, p. 499 – 512, 1976.

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
1992-19	J.G.W. Simons	External Benefits of Transport