

Analysis of State-Independent Importance-Sampling Measures for the Two-Node Tandem Queue

PIETER-TJERK DE BOER
University of Twente

We investigate the simulation of overflow of the total population of a Markovian two-node tandem queue model during a busy cycle, using importance sampling with a state-independent change of measure. We show that the only such change of measure that may possibly result in asymptotically efficient simulation for large overflow levels, is exchanging the arrival rate with the smallest service rate. For this change of measure, we classify the model's parameter space into regions of asymptotic efficiency, exponential growth of the relative error, and infinite variance, using both analytical and numerical techniques.

Categories and Subject Descriptors: G.3 [Probability and Statistics]: *Probabilistic algorithms (including Monte Carlo)*; *Queueing Theory*; I.6.1 [Simulation and Modeling]: Simulation Theory

General Terms: Performance, Theory

Additional Key Words and Phrases: rare-event simulation, importance-sampling, tandem queueing networks

1. INTRODUCTION

Since the end of the 1980s, there has been a steady interest in the problem of estimating rare event probabilities in queueing models, particularly in networks of queues. This interest derives mainly from applications in the field of telecommunications, where buffers in routers and switches should be dimensioned so that overflow is a rare event.

One rare-event simulation method that has received much attention, is *importance sampling* (IS). In this method, the target event is artificially made less rare by changing the probability distributions (also called change of measure or tilting) in the model; this is compensated for by keeping track of the

Note: because of minor copyediting errors in the printed version, this online version is not identical to the printed version.

Author's address: Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands; email: ptdeboer@cs.utwente.nl.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 1515 Broadway, New York, NY 10036 USA, fax: +1 (212) 869-0481, or permissions@acm.org.

© 2006 ACM 1049-3301/06/0700-0225 \$5.00

so-called likelihood ratio. This method can give very large simulation speedup, but the difficulty is finding an appropriate change of measure.

A landmark paper on importance sampling for queueing models is [Parekh and Walrand 1989], in which a rather simple change of measure is proposed for estimating probabilities of overflow in a single queue and of overflow of the total population in networks of queues. For an $M/M/1$ queue, the method boils down to exchanging the arrival rate and the service rate; for such queues in tandem, the arrival rate is exchanged with the rate of the bottleneck server. This change of measure is derived using heuristics based on large-deviations theory. Experimentally, it was shown that the method works well for a single queue, but also that for networks its performance is less consistent, depending strongly on the particular arrival and service rates. Henceforth, this change of measure will be referred to as P&W.

[Sadowsky 1991] proves that the approach from [Parekh and Walrand 1989] is indeed *asymptotically efficient* for single $GI/GI/1$ queues. Asymptotic efficiency means that while the rare event probability decreases exponentially with increasing overflow level, the amount of simulation effort needed for a given relative error (standard deviation divided by the mean) increases less than exponentially with the overflow level.

A less positive result was given in [Glasserman and Kou 1995], where it is shown that for a relatively simple network, namely two (or more) Markovian queues in tandem, the P&W method may or may not be asymptotically efficient. For part of the parameter space (arrival and service rates) a proof was given that the resulting simulation is asymptotically efficient. For another part of the parameter space, a proof was given that the simulation is *not* asymptotically efficient. The remaining part of the parameter space was left undecided.

The P&W change of measure is *state-independent*. This means that the change of measure is a simple change of the arrival and service rates in a way that does not depend on the actual state of the model (e.g., the number of customers in the queues). Use of a *state-dependent* change of measure (i.e., allowing the rates to depend on the numbers of customers in the queue) can produce asymptotic efficiency even in cases where it has been shown that the P&W change of measure is not asymptotically efficient, as demonstrated in [de Boer 2000] and [de Boer and Nicola 2002]. Related approaches are given in [Kollman et al. 1999] and [Ahamed et al. 2006], both of which explicitly try to approximate the change of measure that theoretically gives a zero-variance estimator (which is state-dependent). However, the gain from using a state-dependent change of measure comes at a significant increase in complexity; this extra complexity is mainly due to the large state space of typical models of interest, and becomes even worse when non-Markovian models are considered [de Boer 2005].

So on the one hand, we have P&W's state-independent tilting, which in several cases does not work well; on the other hand, we have the state-dependent tilting which works well even in those cases, but is much more complex. Therefore, even though state-dependent methods are known, it is of interest to bet-

ter understand the exact limitations of state-independent methods. A natural question is whether other state-independent tiltings than P&W could be asymptotically efficient. The fact that P&W is derived on the basis of large-deviations theory, and the fact that an adaptive approach to finding a state-independent tilting fails to result in asymptotic efficiency in some cases (see Section 7.4.4 in [de Boer 2000]), suggest that this might not be the case, but does not constitute a proof.

In Section 3 of this article, we will show that indeed P&W is the only state-independent tilting that can possibly be asymptotically efficient for the two-node tandem queue model. (Whether this result is also true for more general models, is an interesting open question.)

As it turns out, the approach used for this proof can also be used to further analyze the behavior of P&W tilting for the two-node tandem queue. This allows us to pinpoint in Section 4 a larger region in which P&W is not asymptotically efficient than was previously known, and to prove in Section 5 that in part of this region P&W leads to infinite variance.

Finally, for completeness, Section 6 gives an extension of the region in which P&W is asymptotically efficient using a modification of the proof from [Glasserman and Kou 1995], and in Section 7 the performance of P&W tilting is studied numerically.

Clearly, the rest of this article is devoted to estimating one particular overflow probability of just about the simplest nontrivial queueing network. Simulating this particular network by itself is not very interesting. However, the problems that arise when applying P&W to this simple network will presumably also occur in more complicated models. So better simulation techniques will need to be developed, and for doing so, a thorough understanding of the simplest case is helpful. This is discussed in more detail in the concluding Section 8.

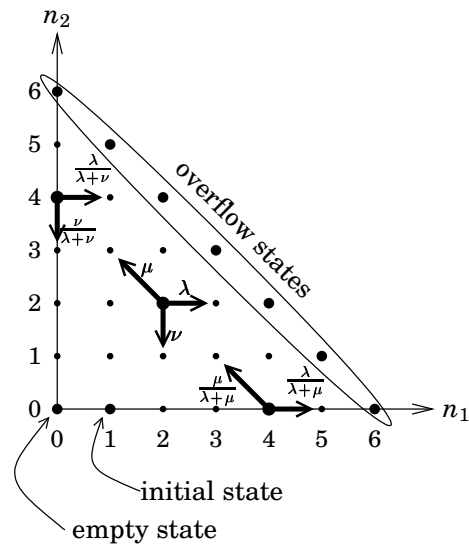
2. PRELIMINARIES

2.1 The Model

This article deals with two $M/M/1$ queues in tandem. Customers arrive to the first queue according to a Poisson process with rate λ . The service time in the first server is exponentially distributed with rate μ . After service completion at the first queue, the customers enter the second queue, the service time of which is exponentially distributed with rate ν . (Note that this notation is slightly different from [Glasserman and Kou 1995]: their μ_1 and μ_2 are written here as μ and ν .) The (rare-event) probability we are interested in is the probability of reaching a state in which the total population of the two queues is K , starting from a state in which there is one customer in the first queue and zero in the second queue, and before the total system becomes empty again. Note that this starting state is equivalent to starting from the completely empty system, since there is only one way to leave the empty state.

We limit the study to cases where both queues are stable, that is, $\lambda < \mu$ and $\lambda < \nu$; otherwise, the overflow event would not be rare. Furthermore, we limit ourselves to the case where the second server is the bottleneck, i.e., $\mu > \nu$.

Fig. 1. Transition probabilities for two queues in tandem, illustrated for $K = 6$.



As noted in [Glasserman and Kou 1995], this is no essential loss of generality. Still, Section 7 also gives numerical results for the case where the first server is the bottleneck.

Since all distributions involved in the model are exponential, and we are not interested in times, we simplify the model to a discrete-time Markov chain. Obvious variables to label the states are n_1 and n_2 , denoting the number of customers in the first and second queues, respectively (including the customers in service). We will also use another set of labels, i and j , defined as $i = n_1 + n_2$ and $j = n_1$, because the total population (now i) plays a more important role in the further calculations than the populations of the individual queues.

The resulting Markov chain is illustrated in Figure 1, which also shows the transition probabilities. We have chosen $\lambda + \mu + \nu = 1$, without loss of generality; thus the transition probabilities are simply λ , μ , and ν in any state in the interior (i.e. where both $n_1 > 0$ and $n_2 > 0$).

2.2 Importance Sampling Simulation

Importance sampling simulation involves changing the underlying probability distributions of the model; this is called a *change of measure* or simply *tilting*. In the case of a discrete-time Markov chain, this is usually done by changing the transition probabilities. In the present article, only tiltings of the following simple form are considered: replace every λ by λ' , every μ by μ' , and every ν by ν' , for any positive λ' , μ' , and ν' with $\lambda' + \mu' + \nu' = 1$. As noted in the introduction, such a tilting is called *state-independent*, since the replacement rates are the same for every state.

Consider a sample path X of this system, represented by the sequence of states it visits: $X = \{Z_1, Z_2, \dots, Z_\tau\}$, with $Z_1 = (1, 0)$ and stopping time $\tau = \min(t | Z_t = (0, 0) \vee Z_t \in S)$ with $S = \{(n_1, n_2) | n_1 + n_2 = K\}$. Furthermore, define

the overflow indicator $I(X) = 1_{Z_\tau \in S}$, and the likelihood ratio

$$L(X) = \prod_{i=1}^{\tau-1} \frac{q_{Z_i \rightarrow Z_{i+1}}}{q'_{Z_i \rightarrow Z_{i+1}}},$$

where q_{\dots} and q'_{\dots} are the transition probabilities in the original and the tilted system, respectively. Then the overflow probability p_K equals $p_K = \mathbb{E}^* I(X) L(X)$, where \mathbb{E}^* denotes expectation in the tilted system. Using simulation to generate N sample paths X_i in the tilted system, the overflow probability can be estimated by the sample average $\hat{p}_K = \sum_{n=1}^N I(X_n) L(X_n) / N$. (See [Heidelberger 1995] for more details.)

The goal of importance sampling simulation is to make the event of interest less rare, in such a way that its probability can be estimated with a small variance. Typically, one strives for *asymptotic efficiency*, meaning in practice that although the overflow probability decreases exponentially in the overflow level K , the amount of simulation effort needed to estimate it with a constant relative error (i.e., the estimator's standard deviation divided by its mean) increases less than exponentially fast in K . Conversely, for a given number of replications, the relative error increases less than exponentially fast in K .

Define m_K as the second moment of the importance sampling estimator: $m_K = \mathbb{E}^* I^2(X) L^2(X) = \mathbb{E} I(X) L(X)$. Then the estimator's relative error is given by $\sqrt{m_K - p_K^2} / p_K \sqrt{N}$, where N is the number of replications. The estimator is asymptotically efficient if [Glasserman and Kou 1995]

$$\limsup_{K \rightarrow \infty} \frac{\log m_K}{\log p_K} = 2.$$

That same article also provides the following limit: $\lim_{K \rightarrow \infty} \frac{1}{K} \log p_K = \log \frac{\lambda}{\nu}$. Combining these leads to the following criterion for asymptotic efficiency:

$$\lim_{K \rightarrow \infty} \frac{1}{K} \log m_K = 2 \log \frac{\lambda}{\nu}. \quad (1)$$

2.3 Basic Equations for the Second Moment

Define $y_{i,j}$ to be the expectation of the second moment of the estimator in the tilted system, starting from state $n_1 + n_2 = i$, $n_1 = j$. Clearly then $m_K = y_{1,1}$. Note the difference: m_K emphasizes the dependence on the overflow level K and presumes starting from state $i = 1, j = 1$, while $y_{i,j}$ emphasizes the starting state and implicitly assumes the overflow level K . For brevity, we write bold \mathbf{y} for the vector containing $y_{i,j}$ for all i, j .

Next, define $y_{i,j}^{(m)}$ as the contribution to $y_{i,j}$ made by sample paths of at most m steps. Consider a state on the "left" boundary, that is, a state with $j = 0$. By conditioning on the next step of the sample path, one easily finds that

$$y_{i,0}^{(m)} = q_{(i,0) \rightarrow (i+1,1)} \frac{q_{(i,0) \rightarrow (i+1,1)}}{q'_{(i,0) \rightarrow (i+1,1)}} y_{i+1,1}^{(m-1)} + q_{(i,0) \rightarrow (i-1,0)} \frac{q_{(i,0) \rightarrow (i-1,0)}}{q'_{(i,0) \rightarrow (i-1,0)}} y_{i-1,0}^{(m-1)}.$$

Substituting the transition probabilities from Figure 1, and repeating the calculation for the other boundary and the "interior", we find the following equa-

tions, for $1 \leq i \leq K - 1$ and $0 \leq j \leq i$:

$$y_{i,0}^{(m)} = \frac{\lambda' + \nu'}{(\lambda + \nu)^2} \left(\frac{\lambda^2}{\lambda'} y_{i+1,1}^{(m-1)} + \frac{\nu^2}{\nu'} y_{i-1,0}^{(m-1)} \right) \quad (2a)$$

$$y_{i,i}^{(m)} = \frac{\lambda' + \mu'}{(\lambda + \mu)^2} \left(\frac{\lambda^2}{\lambda'} y_{i+1,i+1}^{(m-1)} + \frac{\mu^2}{\mu'} y_{i,i-1}^{(m-1)} \right) \quad (2b)$$

$$y_{i,j}^{(m)} = \frac{\lambda^2}{\lambda'} y_{i+1,j+1}^{(m-1)} + \frac{\mu^2}{\mu'} y_{i,j-1}^{(m-1)} + \frac{\nu^2}{\nu'} y_{i-1,j}^{(m-1)} \quad (2c)$$

with boundary conditions, for $0 \leq j \leq K$:

$$y_{0,0}^{(m)} = 0 \quad (2d)$$

$$y_{K,j}^{(m)} = 1 \quad (2e)$$

An appropriate starting vector for this iteration scheme is $\mathbf{y}^{(-1)} = \mathbf{0}$. Then for all $m \geq 0$, the $y_{i,j}^{(m)}$ can be interpreted as the contribution to the second moment of the overflow probability estimator made by sample paths of at most m steps. Hence, the sequence $y_{i,j}^{(m)}$ is monotonically nondecreasing in m , for any i and j . One easily sees that there are only two possibilities:

- either the sequence $\mathbf{y}^{(m)}$ converges to some finite vector $\tilde{\mathbf{y}}$: in that case the estimator's second moment m_K (and thus its variance) is finite and given by $\tilde{y}_{1,1}$, and $\tilde{\mathbf{y}}$ is a stationary, positive, finite solution to (2a)—(2e);
- all elements (except for the boundaries (2d) and (2e)) of the vectors $\mathbf{y}^{(m)}$ run off to infinity as $m \rightarrow \infty$: in that case the variance is infinite, and no stationary, positive, finite solution to (2a)—(2e) exists.

In the sequel we will not use the iterative scheme by doing multiple iterations starting from 0, but by doing a single iteration from some suitable (nonzero) initial $\mathbf{y}^{(0)}$. From this, conclusions can be drawn about the stationary solution of the iteration scheme, and thus about the second moment of the estimator. (E.g., if we prove that no finite positive stationary solution exists, the finite-variance case is impossible, leaving only the infinite-variance case.)

2.4 A Theorem about Iterations

Definition 2.1. For a given matrix $\alpha_{i,j}$, a set J , and an index k_0 , we say that J is *reachable* from k_0 if there exists a sequence of indices $k_1, k_2, k_3, \dots, k_q$ such that $k_q \in J$ and $\alpha_{k_i, k_{i+1}} > 0$ for all $i \in \{0, 1, \dots, q-1\}$.

THEOREM 2.2. *Consider the following iterative scheme:*

$$x_i^{(m)} = \begin{cases} 1 & \text{for } i \in J \\ \sum_j \alpha_{i,j} x_j^{(m-1)} & \text{for } i \notin J \end{cases} \quad (3)$$

for some nonempty set J , and with all $\alpha_{i,j}$ positive and such that J is reachable from any i . Furthermore, we are given some initial vector $x^{(0)}$ such that $0 \leq x_i^{(0)} \leq x_i^{(1)}$ for all i , with $x^{(1)}$ calculated from $x^{(0)}$ by (3).

Then if a finite, positive, stationary vector \tilde{x} for this iteration scheme exists, all of its components are larger than or equal to those of the initial vector: $\tilde{x}_i \geq x_i^{(0)}$ for all i .

PROOF. Assume that the theorem is *not* true; that is, there is a finite, positive, stationary vector \tilde{x} for this iteration scheme, and at least one component of our initial vector $x^{(0)}$ is strictly larger than the corresponding component of the stationary vector.

Define the quantities $f_i = x_i^{(0)}/\tilde{x}_i$ for all i . For $i \notin J$, we find

$$f_i = \frac{x_i^{(0)}}{\tilde{x}_i} \leq \frac{x_i^{(1)}}{\tilde{x}_i} = \frac{\sum_j \alpha_{i,j} x_j^{(0)}}{\sum_j \alpha_{i,j} \tilde{x}_j} = \frac{\sum_j \alpha_{i,j} f_j \tilde{x}_j}{\sum_j \alpha_{i,j} \tilde{x}_j}.$$

If $i \in J$, then $f_i \leq x_i^{(1)}/\tilde{x}_i = 1/1 = 1$. Thus, for each i :

$$f_i \begin{cases} \leq 1 & \text{if } i \in J \\ = f_j & \text{if } i \notin J \text{ and } f_j \text{ equal for all } j \text{ for which } \alpha_{i,j} > 0 \\ < \max_{j:\alpha_{i,j}>0} f_j & \text{otherwise} \end{cases} \quad (4)$$

Set $k_0 = \arg \max_i f_i$, then due to our assumption that the theorem is not true, $f_{k_0} > 1$ and $k_0 \notin J$. Complete the sequence k_j for $j = 1 \dots q$ as given in Definition 2.1. Applying (4) to f_{k_0} , one sees that $f_{k_1} = f_{k_0}$, since we had chosen k_0 such that $f_{k_0} \geq f_j$ for all j . This reasoning can be repeated, leading to $f_{k_q} = f_{k_0} > 1$. However, since $k_q \in J$, we also have $f_{k_q} \leq 1$: a contradiction. Thus, our assumption that the theorem is *not* true, must be incorrect. \square

In Sections 3, 4, and 5, this theorem will be applied to the iterative scheme given by (2a)—(2e) (with an initial vector to be specified later). One easily verifies that indeed the theorem's conditions regarding $\alpha_{i,j}$ are satisfied, noting that the single index i in the theorem corresponds to a *pair* of indices in (2a)—(2e), and with J as the set of overflow states. Note that boundary condition (2d), which is of a form not explicitly allowed by the theorem, can be eliminated by substituting it into (2a)—(2c).

In the sequel, we will refer to the condition $y_{i,j}^{(1)} \geq y_{i,j}^{(0)}$ for all i, j as “growth under iteration”.

2.5 The big O symbol

Throughout this article, we interpret the notation $\mathcal{O}(\epsilon^\alpha)$ for any $\alpha > 0$ as meaning that $\forall \lambda, \mu, \nu : \exists M > 0 : \forall \epsilon, i, j, K : |\mathcal{O}(\epsilon^\alpha)| < M\epsilon^\alpha$. Informally, this means that the bound may depend on λ, μ , and ν , but not on i, j , and K .

3. ONLY P&W CAN BE ASYMPTOTICALLY EFFICIENT

THEOREM 3.1. *For the two-node tandem queue simulation problem defined in Section 2, every state-independent change of measure for which $\lambda' \neq \nu$, or $\mu' \neq \mu$, or $\nu' \neq \lambda$, is not asymptotically efficient. That is, the P&W tilting is the only one that may be asymptotically efficient.*

PROOF. The proof consists of first proposing an initial guess $y^{(0)}$, showing that this guess grows under the iterations if a non-P&W tilting is used, and concluding from this that asymptotic efficiency is not possible.

— Initial Guess

Define the following:

$$y_{i,j}^{(0)} = \left(\frac{\nu^2(1-\epsilon^2)}{\lambda^2} \right)^{i-K} \frac{a_i}{a_K} b_j \quad (5)$$

with

$$a_i = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } i = 1 \\ a_{i-1} \left(1 - \left(\frac{1-2\epsilon}{1-\epsilon} \right)^{i-2} (1-\epsilon) \right)^{-1} & \text{if } i > 1 \end{cases} \quad (6)$$

and

$$b_j = \begin{cases} 1 - \sqrt{\epsilon} & \text{if } j = 0 \\ 1 & \text{otherwise} \end{cases}.$$

The quantity ϵ is a small positive number (e.g., $0 < \epsilon < 1/2$), independent of K , i , and j , which is chosen as small as necessary to satisfy the conditions that show up later.

Note the following properties of a_i , for $i \geq 1$:

$$1 < \frac{a_{i+1}}{a_i} \leq \frac{1}{\epsilon} \quad (7a)$$

$$\frac{a_{i-1}}{a_i} = \frac{a_i}{a_{i+1}} + \mathcal{O}(\epsilon) \quad (7b)$$

$$\{a_i\} \text{ is bounded.} \quad (7c)$$

The proof for (7c) starts by noting that for any fixed $\epsilon \in (0, 1/2)$ and for all sufficiently large i , $\log \left(1 - \left(\frac{1-2\epsilon}{1-\epsilon} \right)^{i-2} (1-\epsilon) \right)^{-1} < 2 \left(\frac{1-2\epsilon}{1-\epsilon} \right)^{i-2} (1-\epsilon)$, using the Taylor expansion for $\log((1-x)^{-1})$. Since $\log a_i$ is a sum of these terms, which decay exponentially in i , it must be upper bounded.

— Growth Under Iteration

Consider one iteration of (2a)—(2e).

For $0 < j < i < K$ (i.e., away from the state-space boundaries), (2c) is used:

$$\begin{aligned} & y_{i,j}^{(1)} \\ &= \left(\frac{\nu^2(1-\epsilon^2)}{\lambda^2} \right)^{i-K} \frac{a_i}{a_K} \left[\frac{\lambda^2 \nu^2}{\lambda' \lambda^2} (1-\epsilon^2) \frac{a_{i+1}}{a_i} b_{j+1} + \frac{\mu^2}{\mu'} b_{j-1} + \frac{\nu^2 \lambda^2}{\nu' \nu^2} \frac{1}{1-\epsilon^2} \frac{a_{i-1}}{a_i} b_j \right] \\ &= y_{i,j}^{(0)} \left[\frac{\nu^2}{\lambda'} (1-\epsilon^2) \frac{a_{i+1}}{a_i} + \frac{\mu^2}{\mu'} (1 - \mathcal{O}(\sqrt{\epsilon})) + \frac{\lambda^2}{\nu'} (1 + \mathcal{O}(\epsilon^2)) \left(\left(\frac{a_{i+1}}{a_i} \right)^{-1} + \mathcal{O}(\epsilon) \right) \right] \\ &= y_{i,j}^{(0)} \left[\frac{\nu^2 a_{i+1}}{\lambda' a_i} + \frac{\mu^2}{\mu'} + \frac{\lambda^2}{\nu'} \left(\frac{a_{i+1}}{a_i} \right)^{-1} + \mathcal{O}(\sqrt{\epsilon}) \right]. \end{aligned}$$

According to Lemma A.1 from the appendix, the factor $\frac{\nu^2}{\lambda'}x + \frac{\mu^2}{\mu'} + \frac{\lambda^2}{\nu'}x^{-1}$ is ≥ 1 for P&W tilting, and > 1 otherwise, for any $x \geq 1$. Thus, for any non-P&W tilting, and for all sufficiently small positive ϵ , we have established that $y_{i,j}^{(1)} > y_{i,j}^{(0)}$ for all $i \geq 1$ and $0 < j < i$.

Next, consider the bottom boundary ($j = i$), where iterations are calculated using (2b):

$$\begin{aligned} y_{i,i}^{(1)} &= y_{i,i}^{(0)} \frac{\lambda' + \mu'}{(\lambda + \mu)^2} \left(\frac{\nu^2}{\lambda'}(1 - \epsilon^2) \frac{a_{i+1}}{a_i} + \frac{\mu^2}{\mu'} (1 - \mathcal{O}(\sqrt{\epsilon})) \right) \\ &\geq y_{i,i}^{(0)} \frac{\lambda' + \mu'}{(\lambda + \mu)^2} \left(\frac{\nu^2}{\lambda'} + \frac{\mu^2}{\mu'} + \mathcal{O}(\sqrt{\epsilon}) \right) \\ &\geq y_{i,i}^{(0)} \left[\left(\frac{\nu + \mu}{\lambda + \mu} \right)^2 + \mathcal{O}(\sqrt{\epsilon}) \right], \end{aligned}$$

where the first \geq uses (7a) and the last step uses Lemma A.2. Since $\nu > \lambda$, it follows that $y_{i,i}^{(1)} > y_{i,i}^{(0)}$ for ϵ sufficiently small.

At the left boundary ($j = 0$), iterations are calculated using (2a):

$$\begin{aligned} y_{i,0}^{(1)} &= y_{i,0}^{(0)} \frac{\lambda' + \nu'}{(\lambda + \nu)^2} \left[\frac{\nu^2}{\lambda'} \frac{a_{i+1}}{a_i} \frac{1}{1 - \sqrt{\epsilon}} + \frac{\lambda^2}{\nu'} \left(\frac{a_{i+1}}{a_i} \right)^{-1} + \mathcal{O}(\epsilon) \right] \\ &= y_{i,0}^{(0)} \frac{\lambda' + \nu'}{(\lambda + \nu)^2} \left[\frac{\nu^2}{\lambda'} x + \frac{\lambda^2}{\nu'} x^{-1} + \frac{\nu^2}{\lambda'} x \sqrt{\epsilon} + \mathcal{O}(\epsilon) \right] \\ &\geq y_{i,0}^{(0)} \left[1 + \frac{\lambda' + \nu'}{(\lambda + \nu)^2} \frac{\nu^2}{\lambda'} \sqrt{\epsilon} + \mathcal{O}(\epsilon) \right], \end{aligned}$$

with $x = \frac{a_{i+1}}{a_i}$ so $x \geq 1$, and where the last step uses Lemma A.3. Clearly, for sufficiently small positive ϵ , we have $y_{i,0}^{(1)} > y_{i,0}^{(0)}$.

Finally, boundary condition (2d) is obviously satisfied, and growth at the other boundary condition (2e) is guaranteed because $y_{K,j}^{(0)} \leq 1$ for all j .

Thus, we have established growth for all i and j .

— Conclusion

Now that we have established that the initial guess (5) grows under the iterations (2a)—(2e) for all i, j , we can conclude from Theorem 2.2 that either a true stationary, positive, finite solution to (2a)—(2e) does not exist, implying infinite variance (cf. Section 2.3); or the true solution $y_{i,j}$ exists but is not less than our initial guess $y_{i,j}^{(0)}$. In the latter case:

$$\begin{aligned} \log m_K &= \log y_{1,1} \geq \log y_{1,1}^{(0)} = \log \left(\left(\frac{\nu^2(1 - \epsilon^2)}{\lambda^2} \right)^{1-K} \frac{a_1}{a_K} \right) \\ &= 2K \log \frac{\lambda}{\nu} - K \log(1 - \epsilon^2) + \log \frac{\nu^2(1 - \epsilon^2)}{\lambda^2} + \log a_1 - \log a_K. \end{aligned}$$

Then

$$\lim_{K \rightarrow \infty} \frac{1}{K} \log m_K \geq 2 \log \frac{\lambda}{\nu} - \log(1 - \epsilon^2) > 2 \log \frac{\lambda}{\nu},$$

which clearly contradicts the necessary condition for asymptotic efficiency (1).

□

Remark 3.2. In the following, an intuitive motivation for the initial guess $y_{i,j}^{(0)}$ is given, which may be of interest for establishing similar proofs for other models.

First, note that for $\epsilon = 0$, $a_i = 1$, and $b_j = 1$, expression (5) gives the second moment a decay rate (in K) that is precisely right for asymptotic efficiency, since the overflow probability itself is proportional to $(\lambda/\nu)^K$. By inserting a small positive ϵ^2 , we make the decay just a bit slower, making asymptotic efficiency impossible. It is not hard to prove that, away from the boundaries, this initial guess leads to growth under the iteration scheme.

However, if $a_i = 1$ for all i , there is a problem at $i = 1$. Recall that $y_{0,0} = 0$ due to the boundary condition, and inserting this into the iteration equations for $i = 1$ leads to a situation in which it is hard or impossible to show that $y_{1,j}$ do increase under the iteration scheme. We need a way for $y_{i,j}$ to go to 0 more smoothly at small i , in order not to disturb the growth-under-iteration property.

The iteration equation for $y_{i+1,j}$ expresses $y_{i+1,j}$ as a linear function with positive coefficients in terms of (among others) $y_{i,j}$. Since, as noted above, for $a_i = 1$ growth under iteration is easy to prove, one could slightly reduce $y_{i,j}$ and still have growth (albeit less) for $y_{i+1,j}$. We can do this by choosing a_i slightly smaller than 1. Having done this, take a look at $y_{i,j}$. We have just chosen it a bit lower than originally, which means that (if still $a_\ell = 1$ for $\ell \neq i$) this $y_{i,j}$ experiences even larger growth under iteration than when a_i was still 1. We can exploit this to make $y_{i-1,j}$ smaller, by choosing $a_{i-1} < 1$; in fact, we can choose $a_{i-1} < a_i$. This reasoning continues, leading to progressively smaller a_i for i closer to 0. If it is possible to let a_i become 0 for some i , a smooth connection to the boundary condition can be achieved.

Our initial guess proposes a series of a_i such that a_i/a_K go to 0 as i goes to 0, based on the above idea. The parameter ϵ determines how slowly a_i/a_K go to 0: the smaller ϵ , the slower.

Finally, the factor b_j just ensures that the initial guess is made slightly smaller at the left boundary; this is a trick to ensure growth there (without it, the growth factor could be as small as $1 - \mathcal{O}(\epsilon)$, which could be less than 1).

4. FURTHER ANALYSIS OF P&W – NO ASYMPTOTIC EFFICIENCY

In this section, we only consider the P&W tilting, and find regions of the parameter space (λ, μ, ν) in which even this tilting does not provide asymptotic efficiency, thus extending Proposition 4.1 from [Glasserman and Kou 1995].

THEOREM 4.1. *For the two-node tandem queue simulation problem defined in Section 2, a state-independent change of measure according to P&W (i.e., with $\lambda' = \nu$, $\mu' = \mu$, and $\nu' = \lambda$) is not asymptotically efficient if $\nu > \frac{3-\mu}{2} - \frac{1}{2}\sqrt{-3\mu^2 - 2\mu + 5}$.*

PROOF. The proof uses an approach similar to the one used for Theorem 3.1

in the previous section: proposing an initial guess, showing that it grows under iteration in the parameter region mentioned in the theorem, and concluding that this makes asymptotic efficiency impossible.

— Initial Guess

First, define δ as an integer ≥ 2 such that

$$\nu - \frac{\mu + \lambda\delta}{1 + \delta} > 0; \quad (8)$$

note that such a δ always exists since the left-hand side goes to $\nu - \lambda$ as $\delta \rightarrow \infty$. Then the initial guess to be used for the proof is the following:

$$y_{i,j}^{(0)} = \left(\frac{\nu^2}{\lambda^2} (1 - \sqrt{\epsilon}) \right)^{i-K} \left(\frac{\lambda}{\mu} \right)^{i-j} \frac{a_{i+\delta j}}{a_{K+\delta K}},$$

with $a_{i+\delta j}$ and $a_{K+\delta K}$ defined by (6). From the properties of a_i listed in (7a)–(7c), it follows for $i > 0$ that

$$1 < \frac{a_{i+b}}{a_i} \leq \frac{1}{\epsilon^b}$$

and

$$\frac{a_{i-c}}{a_i} = \left(\frac{a_{i+b}}{a_i} \right)^{-c/b} + \mathcal{O}(\epsilon)$$

for any fixed (i.e., independent of ϵ, i, j, K) positive integers b and c , with $c \leq i$.

— Growth Under Iterations

In the interior, apply (2a):

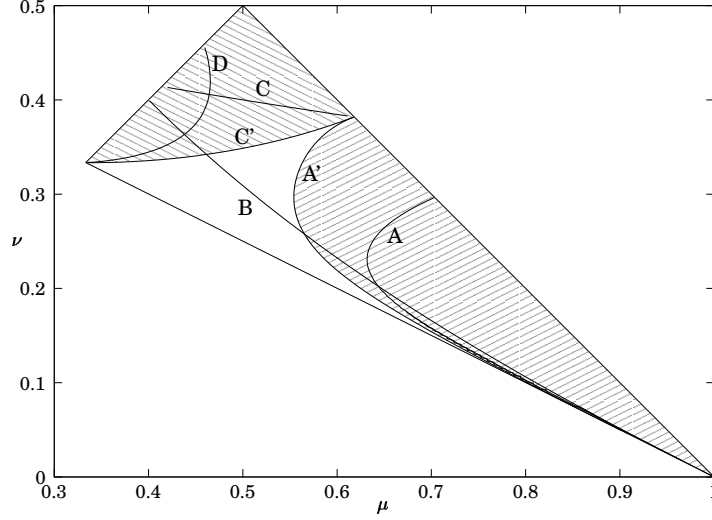
$$\begin{aligned} y_{i,j}^{(1)} &= y_{i,j}^{(0)} \left[\nu(1 - \sqrt{\epsilon}) \frac{a_{i+\delta j+1+\delta}}{a_{i+\delta j}} + \lambda \frac{a_{i+\delta j-\delta}}{a_{i+\delta j}} + \mu(1 - \sqrt{\epsilon})^{-1} \frac{a_{i+\delta j-1}}{a_{i+\delta j}} \right] \\ &= y_{i,j}^{(0)} \left[\nu(1 - \sqrt{\epsilon})x + \lambda \left(x^{-\frac{\delta}{1+\delta}} + \mathcal{O}(\epsilon) \right) + \mu \left(1 + \sqrt{\epsilon} + \mathcal{O}(\epsilon) \right) \left(x^{-\frac{1}{1+\delta}} + \mathcal{O}(\epsilon) \right) \right] \\ &= y_{i,j}^{(0)} \left[\nu x + \lambda x^{-\frac{\delta}{1+\delta}} + \mu x^{-\frac{1}{1+\delta}} + \sqrt{\epsilon} \left(-\nu x + \mu x^{-\frac{1}{1+\delta}} \right) + \mathcal{O}(\epsilon) \right] \end{aligned}$$

where $x = \frac{a_{i+\delta j+1+\delta}}{a_{i+\delta j}}$, so $x > 1$. In order to prove growth, we need to show that the factor between square brackets is larger than 1. For ϵ sufficiently small, the partial derivative of this factor w.r.t. x is positive for all $x \geq 1$ due to (8); therefore, the factor in square brackets is always $\geq 1 + \sqrt{\epsilon}(-\nu + \mu) + \mathcal{O}(\epsilon)$, which clearly is > 1 for any sufficiently small (but nonzero) ϵ because $\mu > \nu$.

At the left boundary, use (2a):

$$\begin{aligned} y_{i,0}^{(1)} &= y_{i,0}^{(0)} \frac{1}{\lambda + \nu} \left[\nu(1 - \sqrt{\epsilon}) \frac{a_{i+1+\delta}}{a_i} + \mu(1 - \sqrt{\epsilon})^{-1} \frac{a_{i-1}}{a_i} \right] \\ &= y_{i,0}^{(0)} \frac{1}{\lambda + \nu} \left[\nu x + \mu x^{-\frac{1}{1+\delta}} + \sqrt{\epsilon} \left(-\nu x + \mu x^{-\frac{1}{1+\delta}} \right) + \mathcal{O}(\epsilon) \right] \end{aligned}$$

where $x = \frac{a_{i+1+\delta}}{a_i}$, so $x \geq 1$. Again, the first derivative w.r.t. x of the factor in square brackets is positive for δ as given by (8), and sufficiently small ϵ and all $x \geq 1$. For $x = 1$, the factor in square brackets is $> \lambda + \nu$ for sufficiently small ϵ , so growth is proven.



- Right of A: P&W is asymptotically efficient according to [Glasserman and Kou 1995].
- Right of A': P&W is asymptotically efficient according to Theorem 6.6.
- Above B: Asymptotic efficiency implies bounded relative error [Glasserman and Kou 1995].
- Above C: P&W is not asymptotically efficient according to [Glasserman and Kou 1995].
- Above C': P&W is not asymptotically efficient according to Theorem 4.1; and according to Theorem 3.1, no state-independent tilting is asymptotically efficient here.
- Left of D: P&W gives infinite variance according to Theorem 5.1.

Fig. 2. Analytically found boundaries in μ, ν space.

At the bottom boundary, (2b) applies:

$$\begin{aligned} y_{i,i}^{(1)} &= y_{i,i}^{(0)} \frac{\nu + \mu}{(\lambda + \mu)^2} \left[\nu(1 - \sqrt{\epsilon}) \frac{a_{i+\delta i+1+\delta}}{a_{i+\delta i}} + \lambda \frac{a_{i+\delta i-\delta}}{a_{i+\delta i}} \right] \\ &= y_{i,i}^{(0)} \frac{\nu + \mu}{(\lambda + \mu)^2} \left[\nu x + \lambda x^{-\frac{\delta}{1+\delta}} + \sqrt{\epsilon}(-\nu x) + \mathcal{O}(\epsilon) \right] \end{aligned}$$

where $x = \frac{a_{i+\delta i+1+\delta}}{a_{i+\delta i}}$, so $x \geq 1$. Again, the first derivative w.r.t. x of the factor in square brackets is positive for sufficiently small ϵ and all $x \geq 1$. Thus, growth is guaranteed for all $x \geq 1$ and sufficiently small ϵ , if $\frac{\nu + \mu}{(\lambda + \mu)^2}(\nu + \lambda) > 1$. Substituting $\nu + \lambda = 1 - \mu$ and rearranging makes this a quadratic equation in ν , which can be solved for ν leading to

$$\nu > \frac{3 - \mu}{2} - \frac{1}{2} \sqrt{-3\mu^2 - 2\mu + 5}.$$

This is precisely the condition stated in the theorem.

Finally, at the diagonal boundary we have $y_{K,j}^{(0)} = (\lambda/\mu)^{K-j} a_{K+\delta j} / a_{K+\delta K} \leq 1$, using $j \leq K$. Comparing with (2e), we see that growth is also guaranteed at the diagonal boundary.

— Conclusion

As in the proof of Theorem 3.1, it follows from the above that either the vari-

ance is infinite, or the estimator's second moment is lower-bounded as follows:

$$\lim_{K \rightarrow \infty} \frac{1}{K} m_K \geq \lim_{K \rightarrow \infty} \frac{1}{K} \log y_{1,1}^{(0)} = 2 \log \frac{\lambda}{\nu} - \log(1 - \sqrt{\epsilon}) > 2 \log \frac{\lambda}{\nu};$$

hence, the resulting simulation indeed is not asymptotically efficient.

□

Figure 2 illustrates the result of Theorem 4.1. The large triangle in this figure represents the ranges of the parameters μ (horizontal axis) and ν (vertical axis) that satisfy the conditions from Section 2 (i.e., both queues stable, and second server is the bottleneck). The area above line C' is the area in which P&W is not asymptotically efficient according to Theorem 4.1; note that Theorem 3.1 implies that *no* state-independent tilting can be asymptotically efficient there. Clearly, in this area $\mu \approx \nu$; this corresponds nicely with earlier observations that P&W does not work well when the service rates are roughly equal.

The other lines in the figure illustrate Theorems 5.1 and 6.6 proved later in this article, and results from [Glasserman and Kou 1995] for comparison.

5. FURTHER ANALYSIS OF P&W – INFINITE VARIANCE

The following theorem specifies a region of the parameter space in which P&W leads to an infinite-variance estimator as $K \rightarrow \infty$. Note that for any finite K , the variance in parts or all of this region is still finite; see also Section 7.3 and Figure 4.

THEOREM 5.1. *For the two-node tandem queue simulation problem defined in Section 2, a state-independent change of measure according to P&W (i.e., with $\lambda' = \nu$, $\mu' = \mu$, and $\nu' = \lambda$) leads to infinite variance if*

$$M^2 < \frac{4\nu\mu\lambda}{1-M}, \quad (9)$$

where M is defined as

$$M = \frac{(\lambda + \mu)^2}{\mu + \nu}. \quad (10)$$

PROOF. The proof for this theorem uses the same approach as the previous two proofs (initial guess, establishing growth, conclusion). However, there is a significant difference in the form of the initial guess. Previous initial guesses were such that at the boundary corresponding to the overflow ($i = K$), they were nonzero. The present initial guess is such that $y_{K,j}^{(0)} = 0$ for all j . Clearly, after one iteration, we will see $y_{K,j}^{(1)} = 1$ due to (2e), so proving growth under iteration at this boundary is no problem. The important consequence however, is that if this initial guess is such that it provides growth under iteration everywhere, then we can multiply it by an arbitrary (positive) factor, and obtain a new, arbitrarily large, initial guess which also grows under iteration. From this it follows (see below) that the true $y_{i,j}$ cannot be upper bounded, and that the variance is infinite.

— Initial Guess

Define a_i and b_i by

$$a_i = \begin{cases} 0 & \text{if } i = 0 \\ a_{i+1}i\epsilon & \text{if } 1 \leq i < 1/\epsilon \\ 1 & \text{if } i \geq 1/\epsilon \end{cases} \quad b_i = \begin{cases} 0 & \text{if } i = 0 \\ b_{i+1}i\epsilon^2 & \text{if } 1 \leq i < 1/\epsilon^2 \\ 1 & \text{if } i \geq 1/\epsilon^2 \end{cases}. \quad (11)$$

The purpose of these a_i and b_i is the same as that of the a_i defined in (6), and for small i , the new a_i is almost equal to the old a_i except for a constant factor. The difference is that the new a_i equal 1 for $i \geq 1/\epsilon$, whereas the earlier a_i never reach their limit value as $i \rightarrow \infty$. The new a_i and b_i have the following properties, for $i \geq 1$, which are rather similar to those of the old a_i :

$$1 \leq \frac{a_{i+1}}{a_i} \leq \frac{1}{\epsilon} \quad 1 \leq \frac{b_{i+1}}{b_i} \leq \frac{1}{\epsilon^2}$$

$$\frac{a_{i-1}}{a_i} = \frac{a_i}{a_{i+1}} + \mathcal{O}(\epsilon) \quad \frac{b_{i-1}}{b_i} = \frac{b_i}{b_{i+1}} + \mathcal{O}(\epsilon^2)$$

Now we consider the following family of initial guesses, parameterized by the positive constant ρ :

$$y_{i,j}^{(0)} = \rho \left(\frac{\nu^2}{\lambda^2} \cdot \frac{\lambda}{1-M} \right)^i \left(\frac{\mu}{M} \right)^j a_j b_{K-i}. \quad (12)$$

Note that this initial guess $y_{i,j}^{(0)}$ is zero both at $j = 0$ (which includes the origin at $i = 0, j = 0$) and at $i = K$.

In the sequel, we assume $K > 1/\epsilon + 1/\epsilon^2$; this is not really a restriction, since we are interested in the asymptotic behavior for large K . Due to this, we have for any i that either $a_{i-1} = a_i = a_{i+1} = 1$ or $b_{K-i-1} = b_{K-i} = 1$ (or both), as can easily be verified.

— Growth

Next, we show that substitution of the initial guess into the iteration equations leads to growth.

The interior: apply (2c) for $0 < i < K$ and $0 < j < K$:

$$\begin{aligned} & y_{i,j}^{(1)} \\ &= y_{i,j}^{(0)} \left(\nu \frac{\lambda}{1-M} \frac{\mu}{M} \frac{a_{j+1}}{a_j} \frac{b_{K-i-1}}{b_{K-i}} + \mu \left(\frac{\mu}{M} \right)^{-1} \frac{a_{j-1}}{a_j} + \lambda \left(\frac{\lambda}{1-M} \right)^{-1} \frac{b_{K-i+1}}{b_{K-i}} \right) \\ &= y_{i,j}^{(0)} \left(\frac{\lambda\mu\nu}{M(1-M)} \left(\frac{b_{K-i+1}}{b_{K-i}} \right)^{-1} \frac{a_{j+1}}{a_j} + M \left(\frac{a_{j+1}}{a_j} \right)^{-1} + (1-M) \frac{b_{K-i+1}}{b_{K-i}} + \mathcal{O}(\epsilon) \right) \end{aligned}$$

For guaranteeing that $y_{i,j}^{(k)}$ increases in the iterations for all sufficiently small ϵ , it is sufficient if

$$\frac{\lambda\mu\nu}{(1-M)M} xy + My^{-1} + (1-M)x^{-1} \geq 1 + c \quad (13)$$

for all $0 < x \leq 1$ and all $y \geq 1$, and for some positive constant c . To prove that the condition (13) is indeed satisfied, start by rewriting its left-hand side:

$$\begin{aligned} \frac{\lambda\mu\nu}{M(1-M)}xy + My^{-1} + (1-M)x^{-1} &= \\ &= \left(\frac{\lambda\mu\nu}{M(1-M)} - \frac{M}{4} \right) xy + \frac{M}{4}xy + My^{-1} + (1-M)x^{-1} \\ &\geq Cx + Mx^{1/2} + (1-M)x^{-1}, \end{aligned}$$

where the last step follows from straightforward analysis of the function $xy/4 + 1/y$, and C is defined as

$$C = \frac{\lambda\mu\nu}{M(1-M)} - \frac{M}{4},$$

so $C > 0$ by (9).

Consider $Mx^{1/2} + (1-M)x^{-1}$ as a function of x : for $x = 1$ it equals 1, while its first derivative w.r.t. x is negative for $x < 1$ if $M \leq 2/3$. Thus we find

$$Cx + Mx^{1/2} + (1-M)x^{-1} \geq \begin{cases} M\sqrt{\frac{1}{2}} + 2(1-M) > 1 & \text{if } x < \frac{1}{2}; \\ \frac{C}{2} + 1 & \text{if } x \geq \frac{1}{2}; \end{cases}$$

so (13) follows with $1 + c$ equal to the minimum of the above two expressions. In the above calculation, we have assumed that $M \leq 2/3$. Lemma A.4 formally establishes this everywhere where (9) holds; but graphically, it is obvious by noting that in Figure 2 the area to the left of curve D clearly has $\nu \geq 1/3$, so $M = (1 - \nu)^2 / (1 - \lambda) \leq 1 - \nu \leq 2/3$.

Next, consider the bottom boundary. Substituting our initial guess, (2b) becomes:

$$\begin{aligned} y_{i,i}^{(1)} &= y_{i,i}^{(0)} \frac{1}{M} \left(\frac{\lambda\mu\nu}{M(1-M)} \frac{a_{i+1}}{a_i} \frac{b_{K-i-1}}{b_{K-i}} + M \frac{a_{i-1}}{a_i} \right) \\ &= y_{i,i}^{(0)} \left(\left(\frac{C}{M} + \frac{1}{4} \right) \frac{a_{i+1}}{a_i} \frac{b_{K-i-1}}{b_{K-i}} + \frac{a_{i-1}}{a_i} \right). \end{aligned}$$

As noted before, either $a_{i-1} = a_i = a_{i+1} = 1$ or $b_{K-i-1} = b_{K-i} = 1$ (or both). In the former case, one sees from the above that $y_{i,i}^{(1)} \geq y_{i,i}^{(0)}$. In the latter case, we have:

$$\begin{aligned} y_{i,i}^{(1)} &= y_{i,i}^{(0)} \left(\left(\frac{C}{M} + \frac{1}{4} \right) \frac{a_{i+1}}{a_i} + \left(\frac{a_{i+1}}{a_i} \right)^{-1} + \mathcal{O}(\epsilon) \right) \\ &\geq y_{i,i}^{(0)} \left(\frac{C}{M} + 1 + \mathcal{O}(\epsilon) \right), \end{aligned}$$

where the last step uses the fact that $a_{i+1} \geq a_i$, and the fact that the function $x/4 + 1/x$ has a minimum of 1 at $x = 2$, as follows from straightforward analysis. Since C and M are positive and independent of ϵ , we have $y_{i,i}^{(1)} \geq y_{i,i}^{(0)}$ for ϵ sufficiently small.

Finally, at the left boundary ($j = 0$), the initial guess is 0. Since all coefficients in the iteration equations, as well as all $y_{i,j}^{(0)}$ values, are nonnegative, the

resulting $y_{i,0}^{(1)}$ will also be nonnegative, thus ensuring that $y_{i,0}^{(1)} \geq y_{i,0}^{(0)}$.

— Infinite Variance

Suppose the variance is finite. In that case, some nonzero stationary solution $\tilde{y}_{i,j}$ to (2a)—(2e) must exist. One can choose ρ in (12) such that for example $y_{1,1}^{(0)} > \tilde{y}_{1,1}$. However, by Theorem 2.2 and the growth-under-iteration proven above, this contradicts the assumption that $\tilde{y}_{i,j}$ is the stationary solution. Therefore, no such finite solution $\tilde{y}_{i,j}$ can exist, and the variance must be infinite.

□

6. EXTENSION OF THE SUFFICIENT CONDITION FOR ASYMPTOTIC EFFICIENCY

Theorem 5.7 in [Glasserman and Kou 1995] gives a sufficient condition for asymptotic efficiency of the P&W estimator in the problem we consider in this article. Below, we derive a stronger condition. The proof from [Glasserman and Kou 1995] will henceforth be referred to as the “original proof”; the present proof is heavily based on it.

Before starting the derivation, define the following convenient notation: $Q_{(n_1, n_2)}$ is the probability of ever reaching a state in the set A (which, as in the original proof, is the set of states with $n_1 > 0$ and $n_2 = 0$), starting from state (n_1, n_2) , before the system empties (i.e., reaches state $(0, 0)$). Furthermore, define $R_{(n_1, n_2)}$ as the probability of ever reaching any state with $n_2 = 0$, starting from state (n_1, n_2) . Obviously, $Q_{(n_1, n_2)} \leq R_{(n_1, n_2)}$ for any n_1 and n_2 . Both Q and R are in the P&W-tilted system; that is, the arrival rate is ν , and the service rates are μ at the first and λ at the second queue; thus, the system is unstable and R is not trivially equal to 1.

In the proofs below, the shorthand notation $(*, i)$ is used to identify sets of states with $n_2 = i$ and any value for n_1 .

LEMMA 6.1. *For any $i \geq 0$, it holds that $R_{(i, 2)} \geq R_{(i+1, 2)}$.*

PROOF. Compare two sample paths, one starting in $(i, 2)$ and one starting in $(i+1, 2)$. One easily sees that, given the same arrival and service times, on the second sample path there will never be fewer customers in either of its queues than on the first sample path. Consequently, the probability of ever reaching $(*, 0)$ is not higher for the second sample path. □

LEMMA 6.2.

$$\sum_{i=0}^{\infty} R_{(i, 2)} \left(\frac{\nu}{\mu}\right)^i \left(1 - \frac{\nu}{\mu}\right) = \left(\frac{\lambda}{\nu}\right)^2.$$

PROOF. Noting that $\left(\frac{\nu}{\mu}\right)^i \left(1 - \frac{\nu}{\mu}\right)$ is the stationary probability distribution of the number of customers in the first queue, this is a straightforward extension of Lemmas 5.3 and 5.4 from the original proof. □

LEMMA 6.3. *The following upper bound holds for all $j \geq 1$:*

$$R_{(j,1)} \leq \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\nu}\right)^2 \left(1 - \frac{\nu}{\mu}\right)^{-1} (1 - \nu).$$

PROOF. Consider starting the P&W-tilted two-node tandem queue in state $(j, 1)$, with $j \geq 1$. Clearly, the probability that it leaves the set $(*, 1)$ by going into $(*, 0)$ is $\lambda/(\lambda + \mu)$, while the probability of leaving via $(*, 2)$ is $\mu/(\lambda + \mu)$. Before leaving $(*, 1)$, external arrivals may occur, causing n_1 to increment; the number of such arrivals is geometrically distributed with parameter ν , so the probability distribution of n_1 upon entrance into $(*, 2)$ is given by

$$\Pr(\text{entrance in } (i, 2)) = \begin{cases} \nu^{i-(j-1)}(1 - \nu) & \text{for all } i \geq j - 1 \\ 0 & \text{otherwise} \end{cases}.$$

Denoting this distribution by π^* , we can upper bound the probability of ever returning from $(\pi^*, 2)$ to $(*, 0)$ as follows:

$$\begin{aligned} R_{(\pi^*, 2)} &= \sum_{i \geq j-1} R_{(i,2)} \nu^{i-(j-1)} (1 - \nu) \leq \sum_{i \geq 0} R_{(i,2)} \nu^i (1 - \nu) \\ &\leq \sum_{i \geq 0} R_{(i,2)} \left(\frac{\nu}{\mu}\right)^i (1 - \nu) \left(1 - \frac{\nu}{\mu}\right) \left(1 - \frac{\nu}{\mu}\right)^{-1} \\ &= \left(\frac{\lambda}{\nu}\right)^2 (1 - \nu) \left(1 - \frac{\nu}{\mu}\right)^{-1}. \end{aligned}$$

The first \leq sign uses Lemma 6.1, the second \leq sign uses $\mu \leq 1$, and the last equality uses Lemma 6.2. (Note that this calculation uses the same technique as used in the bounding of p^* in the original proof.)

Combining all of the above completes the proof of the lemma. \square

LEMMA 6.4. *The following upper bound holds for all $j \geq 0$:*

$$Q_{(j,1)} \leq \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\nu}\right)^2 \left(1 - \frac{\nu}{\mu}\right)^{-1} (1 - \nu) = \frac{\lambda}{\lambda + \mu} + \frac{\lambda^2 \mu^2}{(\mu - \nu) \nu^2}.$$

PROOF. For $j \geq 1$, the result follows immediately from Lemma 6.3. (Note that the equals sign uses the fact that $1 - \nu = \lambda + \mu$.)

From state $(0, 1)$, only two immediate transitions are possible: to $(0, 0)$ and to $(1, 1)$. Since $(0, 0)$ is not in A , the transition to this state does not contribute to $Q_{(0,1)}$. Hence, $Q_{(0,1)} \leq Q_{(1,1)}$. \square

LEMMA 6.5. *The following upper bound holds for all $j \geq 1$:*

$$Q_{(j,0)} \leq \frac{\nu}{\mu + \nu} + \frac{\mu}{\mu + \nu} \left(\frac{\lambda}{\lambda + \mu} + \frac{\lambda^2 \mu^2}{(\mu - \nu) \nu^2} \right).$$

PROOF. Straightforward calculation, by noting that the first term represents moving from $(j, 0)$ to $(j + 1, 0)$, that is, staying in A , and the second term represents all paths that first go through $(j - 1, 1)$, using Lemma 6.4. \square

THEOREM 6.6. *For the two-node tandem queue simulation problem defined in Section 2, the P&W change of measure is asymptotically efficient if*

$$\frac{\mu + \nu}{(\lambda + \mu)^2} \left(\nu + \lambda \mu \left(\frac{1}{\lambda + \mu} + \frac{\lambda \mu^2}{(\mu - \nu) \nu^2} \right) \right) < 1.$$

PROOF. Define q as $\frac{\nu}{\mu + \nu} + \frac{\mu}{\mu + \nu} \left(\frac{\lambda}{\lambda + \mu} + \frac{\lambda^2 \mu^2}{(\mu - \nu) \nu^2} \right)$. By Lemma 6.5, the number of visits N to states in A is stochastically bounded by a geometric random variable with parameter q , so $\Pr\{N = n\} \leq \Pr\{N \geq n\} \leq \sum_{i=n}^{\infty} (1 - q)q^i = q^n$. Hence, we can upper bound $\bar{E}[a^{2N}; T_K < T_0]$ (see the original proof for the meaning of this) as follows:

$$\bar{E}[a^{2N}; T_K < T_0] = \sum_{n=1}^{\infty} a^{2n} \Pr\{N = n\} \leq \sum_{n=1}^{\infty} a^{2n} q^n = \frac{a^2 q}{1 - a^2 q}.$$

This bound is finite and uniform in K if $a^2 q < 1$; substituting a and q completes the proof. \square

Remark 6.7. As shown in Figure 2, the sufficient condition for asymptotic efficiency from Theorem 6.6 (line A') is substantially stronger than the one from [Glasserman and Kou 1995] (line A); this is due to two factors. First, the new proof uses a more direct calculation of $\mathbb{E}[a^{2N}]$ based on the stochastic bounding of N by a geometric random variable, rather than just the mean of N ; second, the new proof explicitly considers what happens at $n_2 = 1$ separately.

One might be tempted to take this approach further and consider also $n_2 = 2$ separately, but that is not so easy, due to the boundary at $n_1 = 0$; at $n_2 = 1$, from this boundary state one has a lower probability of ever reaching A (due to the absorbing state $(0,0)$), but at $n_2 > 1$ this is not trivially true.

Remark 6.8. Note that the necessary condition for asymptotic efficiency per Theorem 4.1 and the sufficient condition from Theorem 6.6 become identical at $\lambda = 0$ (i.e., $\mu + \nu = 1$). This meeting point is at $\mu = \frac{1}{2}\sqrt{5} - \frac{1}{2}$ (note that this number is the inverse of, or 1 less than, the classical "golden ratio" $\frac{1}{2}\sqrt{5} + \frac{1}{2}$) and $\nu = \frac{3}{2} - \frac{1}{2}\sqrt{5}$. Furthermore, the slopes of these two curves also become identical here, so the boundaries of the sufficient and necessary areas touch.

7. NUMERICAL STUDY OF P&W

For a relatively simple model like the two-node tandem queue, it is feasible to numerically calculate the expectation of the second moment (and of the first moment, that is, the overflow probability itself) for a given tilting and a given overflow level. By repeating this for several overflow levels and studying the dependence of the resulting variance on the overflow level, one can classify the simulation as asymptotically efficient or not. Furthermore, cases of infinite variance can be identified. Doing this on a grid of points in the μ, ν parameterspace for the P&W tilting allows us to produce a complete picture of the behavior of P&W, also in the areas where neither of the Theorems 4.1, 5.1 and 6.6 apply (i.e., the white area in Figure 2).

7.1 Numerical Calculation of the Relative Error

In order to calculate the relative error, we need to find the expectation of the first and the second moment of the estimator, that is, $p_K = \mathbb{E}^*LI$ and $m_K = \mathbb{E}^*L^2I$ respectively (cf. Section 2.2; for brevity, the dependence of L and I on the sample path X is now implicit), on sample paths that start at $n_1 = 1, n_2 = 0$ and continue until either $n_1 + n_2 = K$ (overflow) or $n_1 = n_2 = 0$ (system empty). Here \mathbb{E}^* denotes the expectation in the tilted system. Since L is the likelihood ratio between the original and the tilted system, this is equivalent to finding $\mathbb{E}L^{k-1}I$ for $k \in \{1, 2\}$.

Consider sample paths starting in the starting state $n_1 = 1, n_2 = 0$, and continuing until either $n_1 + n_2 = i$ or $n_1 = n_2 = 0$. For these sample paths, define $w_{i,j}^{(k)} = \mathbb{E}L^{k-1}I_{i,j}$, where $I_{i,j}$ is the indicator of ending in the state $n_1 + n_2 = i, n_1 = j$. Then clearly $p_K = \sum_{j=0}^K w_{K,j}^{(1)}$ and $m_K = \sum_{j=0}^K w_{K,j}^{(2)}$. Note the difference in perspective between $y_{i,j}$ as defined in Section 2.3, and $w_{i,j}^{(2)}$ defined here: both are concerned with the expectation of the second moment, but the former has an explicit (in the subscript) starting state and an implicit target (overflow) set, while the latter has an implicit starting and an explicit target state. Next we define vectors w_i as follows, omitting the superscript (k) for brevity:

$$w_i = (w_{i,0}, w_{i,1}, w_{i,2}, \dots, w_{i,i}, 0, 0, \dots).$$

It turns out to be quite straightforward to calculate these vectors, iteratively starting from $i = 1$ (which is trivial because it is the initial level) for all i , as shown below. For convenience, we henceforth refer to the index $i = n_1 + n_2$ as the “level”, and to a state with $n_1 + n_2 = i$ and $n_1 = j$ as “state j at level i ”, and we omit the superscript (k).

Define a matrix Q_i in which the ℓ, m element is (for $k = 1$) the probability that starting from state ℓ at level i , level $i+1$ will be reached before level 0, and the first state at level $i+1$ will be state m . For $k = 2$, instead of these probabilities, the expected contributions to the second moment are put into the matrix. Then clearly

$$w_{i+1} = w_i Q_i.$$

Hence,

$$w_i = w_1 Q_1 Q_2 \dots Q_{i-1},$$

where $w_1 = (0, 1, 0, 0, \dots)$, so the problem has been reduced to that of calculating the Q matrices.

Next, we express Q_i in terms of the elementary transition probabilities of the Markov chain and (for $k = 2$) the corresponding contributions to the second moment. These contributions are summarized in three groups of matrices: A_i, B_i , and C_i . The ℓ, m element of A_i is the probability, multiplied by the likelihood ratio to the power $k - 1$, of a transition from state ℓ at level i to state m at level $i+1$ (i.e., an arrival to the first queue). Similarly, B_i represents the transitions in which the level stays the same (i.e., service completion at the first queue), and C_i represents the transitions in which the level decreases by 1 (i.e., service completion at the second queue, after which the customer leaves

the system). One easily finds:

$$Q_i = A_i + B_i Q_i + C_i Q_{i-1} Q_i.$$

This can be solved for Q_i , yielding

$$Q_i = (Id - B_i - C_i Q_{i-1})^{-1} A_i, \quad (14)$$

where Id is the identity matrix. Note that at level $i = 1$, the matrix C_1 is 0 since any transition to the lower level empties the system, which ends the sample path.

Composing the A , B , and C matrices by inspection is straightforward. As an example, this is A_3 for both $k = 1$ and $k = 2$:

$$A_3 = \begin{bmatrix} 0 & \left(\frac{\lambda}{\lambda+\nu}\right)^k \left(\frac{\lambda'+\nu'}{\lambda'}\right)^{k-1} & 0 & 0 & 0 & 0 \dots \\ 0 & 0 & \lambda^k \left(\frac{1}{\lambda'}\right)^{k-1} & 0 & 0 & 0 \dots \\ 0 & 0 & 0 & \lambda^k \left(\frac{1}{\lambda'}\right)^{k-1} & 0 & 0 \dots \\ 0 & 0 & 0 & 0 & \left(\frac{\lambda}{\lambda+\mu}\right)^k \left(\frac{\lambda'+\mu'}{\lambda'}\right)^{k-1} & 0 \dots \\ 0 & 0 & 0 & 0 & 0 & 0 \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Unfortunately, the matrices to be inverted in (14) for the calculation of the second moment under the P&W tilting often turn out to be ill-conditioned. The conditioning can be improved significantly by multiplying the matrix components by numbers of the form α^d , where d is the distance from the diagonal, and α is some constant to be chosen suitably (e.g., iteratively such that the condition number is small enough). One easily sees that the influence of this multiplication can be undone by multiplying the inversion result by α^{-d} .

The numerical results presented later in this section were checked for any remaining numerical errors by repeating some of them at a much higher numerical precision (40 decimal digits). It was found that the results obtained using machine-precision (IEEE double precision, i.e., about 16 decimal digits) agreed to at least 10, but in most cases 12 to 14 decimal places with those found in the (very much slower) high-precision calculation, so apparently the machine precision combined with the above conditioning technique is sufficiently accurate (note though, that due to the CPU and memory requirements, this verification was not feasible for overflow levels higher than 160).

Note: the method for calculating $\mathbb{E}^* L^k I^k$ described in this section is an extension of the methods developed in Appendix A.2 of [Garvels and Kroese 1999] and Chapter 2 of [de Boer 2000].

7.2 Classification

In the analysis in Sections 3 through 6, we have distinguished between three fundamentally different kinds of asymptotic behavior of the importance sampling simulation: asymptotic efficiency (possibly with bounded relative error), exponential (or faster) growth of the relative error, and infinite variance. We

would like to classify the results of numerical calculation of the estimator's variance using the same categories.

Infinite variance is easy to detect: if for some level i (and $k = 2$) the matrix inversion in (14) is not possible, or the resulting inverse is such that it yields negative numbers in the subsequent calculation of w , then apparently no (positive) vector $w_i^{(2)}$ exists, so no (finite, real, positive) values exist that satisfy the equations governing the second moment. (Note that this has only been observed for $k = 2$, which is good, since for $k = 1$ probabilities are calculated, which exist always.)

In order to recognize asymptotic efficiency (or its absence), one needs to calculate the relative error at several overflow levels K and compare them. For comparing the relative errors, the following measure can be used (denoting by RE_K the relative error at overflow level K , defined¹ as $\sqrt{m_K - p_K^2/p_K}$):

$$r(K) = \frac{\log RE_K - \log RE_{K/2}}{\log RE_{K/2} - \log RE_{K/4}}.$$

One easily verifies that if the RE grows exponentially with K , then $r(K) = 2$ for all K . Similarly, if the RE grows polynomially or, for example, logarithmically, then $\lim_{K \rightarrow \infty} r(K) = 1$.

On the other hand, if $\lim_{K \rightarrow \infty} r(K) = 0$, the RE is bounded. Actually, the RE is bounded under a more general condition, namely if $r(K) \leq \epsilon$ for all $K > m$, for some m and some $\epsilon < 1$. To prove this, note that it follows from $r(K) \leq \epsilon$ that $\log RE_K - \log RE_{K/2} \leq \epsilon(\log RE_{K/2} - \log RE_{K/4})$. Repeated application of this leads for any positive integer i to

$$\begin{aligned} \log RE_{m \cdot 2^i} &\leq \log RE_m + (\log RE_m - \log RE_{m/2}) (\epsilon + \epsilon^2 + \epsilon^3 + \cdots + \epsilon^i) \\ &< \frac{1}{1 - \epsilon} \log RE_m - \frac{\epsilon}{1 - \epsilon} \log RE_{m/2}. \end{aligned}$$

Thus, we have² an upper bound for RE_K for large K .

7.3 Results

The result of a scan of the entire μ, ν parameter space for the two-node tandem queue with P&W tilting is shown in Figure 3. Note that, in contrast to other results in this article, this scan also covers the case in which the first queue is the bottleneck (i.e., $\mu < \nu$).

At every point in a grid with a stepsize of 0.0025 (and 0.0008333 in the magnified area), the value of r as defined above was calculated, and indicated in the graph using a shade of gray. Points having $r = 0$ got the darkest gray (labeled as “bounded RE”, although actually any $r < 1$ corresponds to bounded RE), points having $r = 2$ got middle gray (“exponential growth”); points having r values in between got a proportional intermediate shade. Points at which the

¹normalized to 1 replication; cf. Section 2.2 with $N = 1$.

²Strictly speaking, this is not a bound for all K , but only for those that can be written as $m \cdot 2^i$ for all positive integers i and for a fixed integer m , since the bound is expressed in terms of RE_m and $RE_{m/2}$. However, there is little reason to believe that RE_K would not be a monotone function of K .

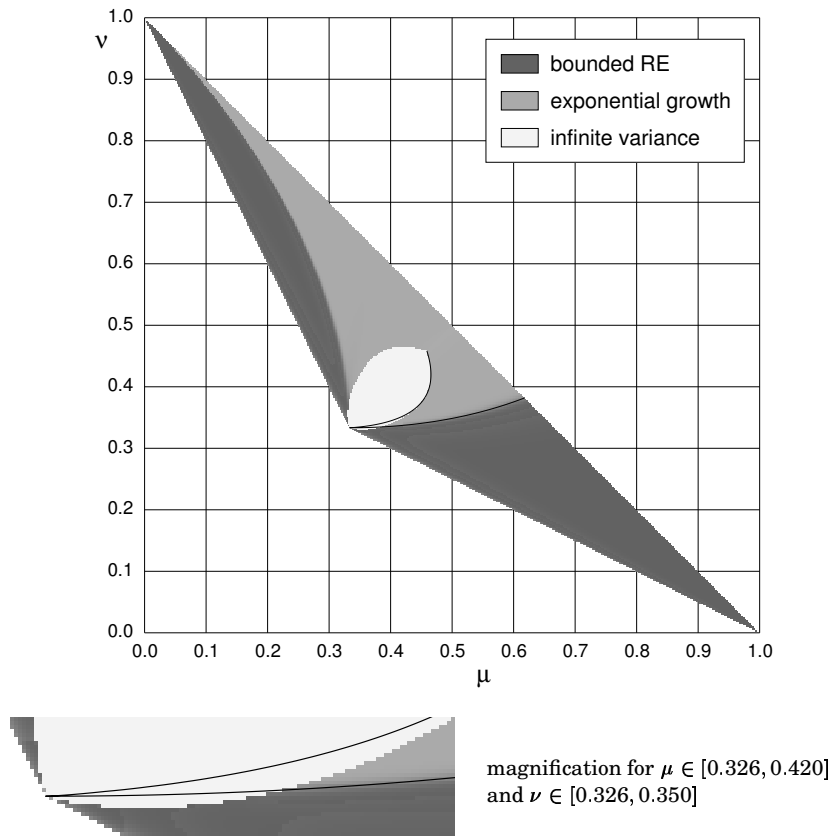


Fig. 3. Behavior of the two-node tandem queue simulation under P&W tilting.

calculation of the second moment was not possible, so for which the variance was apparently infinite as discussed above, got the lightest shade. The overflow level K used for each of these points varies, from $K = 160$ for most points up to $K = 1280$ for a few. The higher overflow levels were used at points where a lower K yielded an r value that was not close to 0 or 2; in such cases r indeed became closer to 0 or 2 (or infinite variance occurred) at higher K , suggesting that the asymptotic behavior was not yet dominant at the lower K .

Theorems 4.1 and 5.1 provide sufficient conditions for exponential growth and infinite variance; these curves (i.e., curves C' and D from Figure 2) have also been drawn in the figure, for comparison with the numerical results. Clearly, these curves are pretty close to the actual boundaries of the respective regions, but there are some points with infinite variance that are outside the areas delineated by the curves (some of these points were verified as discussed at the end of Section 7.1, so it seems very unlikely that this is due to numerical inaccuracies). Thus, the conditions in Theorems 4.1 and 5.1 are indeed only sufficient and not necessary conditions. However, outside the infinite variance area, the curve from Theorem 4.1 does seem to be the true boundary between the areas of asymptotic efficiency and exponential growth, although

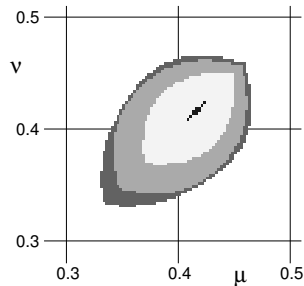


Fig. 4. A closer look at the infinite variance region:
 black points in center: infinite variance for $K = 7$;
 light gray: infinite variance for $K = 10$ but not for $K = 7$;
 medium: infinite variance for $K = 20$ but not for $K = 10$;
 dark: infinite variance for $K = 160$ but not for $K = 20$.

this has not been formally proven.

Furthermore, note that although the proofs only imply bounded relative error above line B in Figure 2, the numerical results suggest that wherever the simulation is asymptotically efficient, it also has a bounded relative error. In other words, $r \approx 1$ was never observed.

It is remarkable that the figure has a partial symmetry under exchange of μ and ν (i.e., making the first or the second queue the bottleneck): the infinite variance region seems to be symmetric, while the asymptotically efficient region clearly is not symmetric. Since the problem is not symmetric under exchange of μ and ν , it is in principle not to be expected that the behavior should be symmetric. One can verify that for every cyclic path sufficiently far away from the absorbing states ($n_1 + n_2 = 0$ and $n_1 + n_2 = K$), a cyclic path with the same probability and likelihood ratio exists in the system obtained by exchanging μ and ν ; and since infinite variance is related to cyclic paths (cf. Lemma 3.2 in [Randhawa and Juneja 2004]), this makes it plausible that the infinite variance region should indeed be symmetric. Furthermore, note that the asymmetry of the exponential growth region confirms the empirical observation of [Parekh and Walrand 1989] that importance sampling simulating with $\nu < \mu$ performs better than with $\nu > \mu$.

Finally, it should be noted that the classification in Figure 3 is purely based on the asymptotic behavior for high overflow levels. Thus, equal shades in the figure do not mean equal simulation performance at a given value of K . For example, infinite variance already occurs for all $K \geq 7$ in the center of the infinite variance region near $\mu = \nu = 0.416$; but the closer to the boundary of the region one gets, the higher K can be without infinite variance, as illustrated in Figure 4. Similarly, the exponential growth rate can be very different among points within the exponential growth area.

8. CONCLUSIONS

In this article, a detailed analysis of simulating a Markovian two-node tandem queue using importance sampling with a state-independent change of measure has been presented. The main conclusion is that the only such change of measure that may be asymptotically efficient, is the one proposed in [Parekh and Walrand 1989]. Furthermore, the performance of this change of measure has been characterized further, thus extending the results from [Glasserman and Kou 1995]. Although these results have been obtained for one particular

model, they do have implications for research in the general area of efficient simulation of queueing models, as follows.

First, the results emphasize the importance of state-dependent tilting. It is now certain that no state-independent tilting can be asymptotically efficient for the two-node tandem model at certain parameter values; given the model's simplicity, this is presumably also true for almost any other nontrivial queueing model. This means that methods for finding a state-independent tilting have limited applicability (although when they do work, they still may be preferable due to their simplicity), which strongly motivates further research into state-dependent tilting.

Second, results from this article are useful for evaluating other importance sampling measures. The complete characterization of P&W for the two-node tandem queue allows one to select the most interesting parameter settings for testing a newly proposed importance sampling measure. Furthermore, the techniques introduced in this article (for proving the absence of asymptotic efficiency and for numerically characterizing an importance sampling estimator's performance) may also be used to characterize other importance sampling measures.

Future extensions of this work could be the following. It would be of interest to see whether a stronger version of Theorem 3.1 could be proved, asserting that P&W is the only possibly asymptotically efficient state-independent change of measure for a larger class of models. Also, analysis of state-independent exponential tilting of a non-Markovian queueing network would be an interesting extension.

APPENDIX

In this appendix, we prove some auxiliary lemmas; these give upper or lower bounds on some functions of the original (λ , μ , and ν) and tilted (λ' , μ' , and ν') rates. It is assumed that these rates satisfy the conditions listed in Sections 2.1 and 2.2.

LEMMA A.1. *The following expression holds for all $x \geq 1$:*

$$\frac{\nu^2}{\lambda'}x + \frac{\mu^2}{\mu'} + \frac{\lambda^2}{\nu'}x^{-1} \geq 1,$$

with strict equality only possible (but not necessary) for P&W tilting.

PROOF. Apply the Cauchy-Schwarz inequality to the two vectors

$$\left(\frac{\nu}{\sqrt{\lambda'}}\sqrt{x}, \frac{\mu}{\sqrt{\mu'}}, \frac{\lambda}{\sqrt{\nu'}}\sqrt{x^{-1}} \right) \text{ and } \left(\sqrt{\lambda'}, \sqrt{\mu'}, \sqrt{\nu'} \right)$$

to find

$$\left(\frac{\nu^2}{\lambda'}x + \frac{\mu^2}{\mu'} + \frac{\lambda^2}{\nu'}x^{-1} \right) \cdot (\lambda' + \mu' + \nu') \geq (\nu x^{1/2} + \mu + \lambda x^{-1/2})^2.$$

Substitute $\lambda' + \mu' + \nu' = 1$ to obtain

$$\frac{\nu^2}{\lambda'}x + \frac{\mu^2}{\mu'} + \frac{\lambda^2}{\nu'}x^{-1} \geq (\nu x^{1/2} + \mu + \lambda x^{-1/2})^2.$$

One easily sees that the first derivative of the r.h.s. w.r.t. x is > 0 for all $x > 1$, so its minimum (on the domain) is 1 and attained only at $x = 1$. Equality in the Cauchy-Schwarz inequality is only attained when the two vectors are parallel; with $x = 1$ this implies $\lambda' = \nu$, $\mu' = \mu$, and $\nu' = \lambda$; that is, P&W tilting. This completes the proof. \square

LEMMA A.2. *For each allowable (per Section 2) set of rates, the following holds:*

$$(\lambda' + \mu') \left(\frac{\nu^2}{\lambda'} + \frac{\mu^2}{\mu'} \right) \geq (\mu + \nu)^2.$$

PROOF. Calculate the first partial derivative with respect to λ' of the left-hand side:

$$\frac{\partial}{\partial \lambda'} (\lambda' + \mu') \left(\frac{\nu^2}{\lambda'} + \frac{\mu^2}{\mu'} \right) = \frac{\mu^2}{\mu'} - \frac{\nu^2 \mu'}{\lambda'^2} = \frac{\mu^2 \lambda'^2 - \nu^2 \mu'^2}{\mu' \lambda'^2}.$$

This implies that the left-hand side attains its minimum at $\lambda' = \nu \mu' / \mu$, and substitution shows that this minimum is $(\mu + \nu)^2$. \square

LEMMA A.3. *For each allowable (per Section 2) set of rates, the following holds for all $x \geq 1$:*

$$\frac{\lambda' + \nu'}{(\lambda + \nu)^2} \left(\frac{\nu^2}{\lambda'} x + \frac{\lambda^2}{\nu'} x^{-1} \right) \geq 1.$$

PROOF. By setting the derivative with respect to λ' of the left-hand side to zero, one finds that this left-hand side has its minimum at $\lambda' = \nu \nu' x / \lambda$ (this sole extremum cannot be a maximum because the left-hand side goes to infinity as $\lambda' \rightarrow 0$). Substituting this, we calculate

$$\frac{\lambda' + \nu'}{(\lambda + \nu)^2} \left(\frac{\nu^2}{\lambda'} x + \frac{\lambda^2}{\nu'} x^{-1} \right) \geq \frac{\nu'}{(\lambda + \nu)^2} \left(1 + \frac{\nu}{\lambda} x \right) \left(\frac{\nu \lambda}{\nu'} + \frac{\lambda^2}{\nu'} x^{-1} \right) = \frac{(\lambda + \nu x)^2}{(\lambda + \nu)^2 x} \geq 1,$$

where the last step follows from $(\lambda + \nu x)^2 - (\lambda + \nu)^2 x = (\nu^2 x - \lambda^2)(x - 1) \geq 0$ since $\nu \geq \lambda$ and $x \geq 1$. \square

LEMMA A.4. *For each allowable (per Section 2) set of rates, and with M defined as*

$$M = \frac{(\lambda + \mu)^2}{\mu + \nu} = \frac{(1 - \nu)^2}{\mu + \nu} = \frac{(1 - \nu)^2}{1 - \lambda}$$

the following holds: if

$$M^2 - M^3 < 4\lambda\mu\nu \tag{15}$$

then

$$M < 2/3.$$

PROOF. Since $\partial(4\lambda(1 - \lambda - \nu)\nu)/\partial\lambda = 4\nu(1 - \nu - 2\lambda) = 4\nu(\mu - \lambda) \geq 0$, the right-hand side of (15) will decrease monotonically as λ decreases from ν to 0 at constant ν .

Considering $M^2 - M^3$ as a function of M on its domain $[0, 1]$, one sees that it is continuous and has a single maximum at $M = 2/3$; furthermore, one trivially

sees that $\partial M/\partial \lambda > 0$ (at constant ν). Thus, as λ decreases from ν to 0, the left-hand side of (15) will not decrease until λ is small enough to make $M < 2/3$.

At $\lambda = \nu$, condition (15) is not satisfied because $M^2 - M^3 - 4\lambda\mu\nu = (1 - \nu)^2 - (1 - \nu)^3 - 4\nu^2(1 - 2\nu) = \nu(1 - 3\nu)^2 \geq 0$. Thus, it follows that (15) cannot be satisfied until $M < 2/3$. \square

ACKNOWLEDGMENTS

The author thanks the anonymous reviewers for their helpful comments and, in particular, one of the reviewers for simplifying the proof of Lemma A.1.

REFERENCES

- AHAMED, T. P. I., BORKAR, V. S., AND JUNEJA, S. 2006. Adaptive importance sampling technique for Markov chains using stochastic approximation. *Oper. Res.* To appear.
- DE BOER, P. T. 2000. Analysis and efficient simulation of queueing models of telecommunication systems. Ph.D. thesis, University of Twente, The Netherlands.
- DE BOER, P. T. 2004. Analysis of state-independent measures for the two-node tandem queue (extended abstract). In *the 5th International Workshop on Rare Event Simulation and Combinatorial Optimization (RESIM/COP'04)*. Budapest, Hungary.
- DE BOER, P. T. 2005. Rare-event simulation of non-Markovian queueing networks using a state-dependent change of measure determined using cross-entropy. *Ann. Oper. Res.* 134, 1, 69–100.
- DE BOER, P. T., KROESE, D. P., AND RUBINSTEIN, R. Y. 2004. A fast cross-entropy method for estimating buffer overflows in queueing networks. *Manage. Sci.* 50, 7, 883–895.
- DE BOER, P. T. AND NICOLA, V. F. 2002. Adaptive state-dependent importance sampling simulation of markovian queueing networks. *European Trans. on Telecomm.* 13, 4, 303–315.
- GARVELS, M. J. J. AND KROESE, D. P. 1999. On the entrance distribution in RESTART simulation. In *2nd International Workshop on Rare Event Simulation, RESIM'99*. 65–88.
- GLASSERMAN, P. AND KOU, S.-G. 1995. Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Model. Comput. Simul.* 5, 1, 22–42.
- HEIDELBERGER, P. 1995. Fast simulation of rare events in queueing and reliability models. *ACM Trans. Model. Comput. Simul.* 5, 1, 43–85.
- KOLLMAN, C., BAGGERLY, K., COX, D., AND PICARD, R. 1999. Adaptive importance sampling on discrete markov chains. *Ann. Appl. Probab.* 9, 391–412.
- PAREKH, S. AND WALRAND, J. 1989. A quick simulation method for excessive backlogs in networks of queues. *IEEE Trans. Autom. Control* 34, 54–66.
- RANDHAWA, R. S. AND JUNEJA, S. 2004. Combining importance sampling and temporal difference control variates to simulate markov chains. *ACM Trans. Model. Comput. Simul.* 14, 1, 1–30.
- SADOWSKY, J. S. 1991. Large deviations theory and efficient simulation of excessive backlogs in a $GI/GI/m$ queue. *IEEE Trans. Autom. Control* 36, 1383–1394.

Received July 2005; revised March 2006; accepted March 2006