Variance Reduction Applied to Product-Form Multi-Class Queuing Networks

Bruno Tuffin
IRISA

Performance of product-form multi-class queuing networks can be determined from normalization constants. For large models, the evaluation of these performance metrics is not possible because of the required amount of computer resources (either by using normalization constants or by using MVA approaches). Such large models can be evaluated with Monte Carlo summation and integration methods. This paper proposes two cluster sampling Monte Carlo techniques to deal with such models. First, for a particular type of network, we propose a variance reduction technique based on antithetic variates. It leads to an improvement of Ross, Tsang and Wang's algorithm which is designed to analyze the same family of models. Second, for a more general class of models, we use a mixture of Monte Carlo and quasi-Monte Carlo methods to improve the estimate with respect to Monte Carlo alone.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: Probabilistic algorithms (including Monte Carlo); I.6.3 [**Simulation and Modeling**]: Applications; I.6.8 [**Simulation and Modeling**]: Types of Simulation—*Monte Carlo*

General Terms: Theory, Algorithms, Performance

Additional Key Words and Phrases: Product-form networks, Monte Carlo, Variance reduction, antithetic variates, low discrepancy sequences

1. INTRODUCTION

Communication and computer systems can be efficiently represented by stochastic models. Among them, closed product-form multi-class Jackson networks are useful in practice. The steady-state solution to such networks is known, but it includes a normalization constant for which, in general, no closed form is known. Thus an important problem in the area is the development of efficient algorithms (like MVA [Reiser and Kobayashi 1975]) and, in particular, the computation of those normalization constants, which are sums over all the possible states of the network. These calculation methods (MVA, convolution [Buzen 1973], etc) are quickly irrelevant, when the number of states increases. In these cases, we can use approximation methods.

Address: IRISA, Campus de Beaulieu, 35042 Rennes cédex, France, E-mail: tuffin@irisa.fr

In Bell Laboratories, McKenna, Mitra and Ramakrishnan used in the eighties the asymptotic expansion of the normalization constants [McKenna and Mitra 1984][Mckenna and Mitra 1982][Mckenna et al. 1981][Ramakrishnan and Mitra 1982] to derive efficient approximation algorithms. Nevertheless, we think that the best approaches are Monte Carlo ones. Ross *et al.* [Ross et al. 1994][Ross and Wang 1993][Wang and Ross 1994][Ross and Wang 1997] have developed a software package, MonteQueue, which uses four different techniques more or less efficient with respect to the particularities of the studied network. Two of these methods use the normalization constant in a summation form. The two last techniques use the fact that, under some restrictive conditions, the normalization constant can be represented by an integral. All these approaches give good results. We propose here refinements to improve their efficiency, in order to take less time to obtain the same confidence interval width. For one of the methods working on the integral form, we propose a variance reduction algorithm based on antithetic variates and we prove that it improves the corresponding technique, described in [Ross et al. 1994].

Another general approach for numerical integration or summation is the use of quasi-Monte Carlo. In such methods, the independent and identically distributed random sequence is replaced by a deterministic one, called *low discrepancy sequence*, which is optimally distributed (see [Niederreiter 1992][Bouleau and Lépingle 1993][Niederreiter 1978]). Unfortunately, the error bound has only a theoretical interest because its value is very large for a fixed number of iterations and a large dimension of the integration or summation space. However, it is possible to make a mixture of Monte Carlo and quasi-Monte Carlo methods. As a matter of fact, we can use the distribution of low discrepancy sequences to obtain a variance reduction in Monte Carlo. This idea was introduced by Cranley and Patterson [Cranley and Patterson 1976] in 1976 and Shaw [Shaw 1988] in 1988 for Bayesian statistics. Owen [Owen 1995] [Owen 1994] uses the same type of technique. In some cases, the convergence speed of low discrepancy sequences ensures a variance reduction if one uses a sufficient number of elements of the low discrepancy sequence.

The paper is organized as follows: In Section 2 we present the model of product-form queuing networks and the Monte Carlo methods of Ross *et al.* Section 3 deals with the application of antithetic variates method. In Section 4, we give a short review of quasi-Monte Carlo methods, which we next apply as a variance reduction technique. Finally, we conclude in Section 5.

2. MODEL AND OVERVIEW OF ROSS *ET AL.* WORK

2.1 Model

Our network has M stations, J classes and a population of N_j customers for class j. We suppose that there exist two types of stations: first come first served stations (FCFS) and infinite server stations (IS). We suppose, without loss of generality, that stations 1 to L are FCFS and stations $L+1$ to M are IS, and we denote by s_m the number of servers at FCFS station m ($1 \leq m \leq L$). All classes have the same exponential service for each FCFS station with mean $1/\mu_m$ for $m = 1, \cdots, L$. On the other hand, all classes may have different exponential services in IS stations with mean $1/\mu_{jm}$ for class j in station m.

The routing is supposed to be Markovian. For each class j, let λ_{jm} be the relative

visit ratio of a customer of class j to station m. Let $\rho_{jm} = \lambda_{jm}/\mu_m$ for $m = 1, \cdots, L$ and $\rho_{jm} = \lambda_{jm}/\mu_{jm}$, for $m = L+1, \cdots, M$, be the traffic intensity for class j at station m. We denote by

$$\rho_{j0} = \sum_{m=L+1}^{M} \rho_{jm} \ 1 \leq j \leq J$$

the total traffic intensity for the whole set of IS stations.

A FCFS station is said to be in *normal usage* [Mckenna and Mitra 1982] if

$$\sum_{j=1}^{J} N_j \frac{\rho_{jm}}{\rho_{j0}} < 1,$$

in *critical usage* [Ross and Wang 1993] if

$$\sum_{j=1}^{J} N_j \frac{\rho_{jm}}{\rho_{j0}} = 1,$$

and *near the critical usage* [Ross and Wang 1993] if

$$\sum_{j=1}^{J} N_j \frac{\rho_{jm}}{\rho_{j0}} \approx 1.$$

The number $\sum_{j=1}^{J} N_j \rho_{jm}/\rho_{j0}$ can be asymptotically seen as an indicator of the utilization of the m^{th} station [Mckenna and Mitra 1982]. The network is said to be in normal usage if all the FCFS stations are in normal usage. It is said to be in mixed usage if all the stations are in normal usage or near critical usage. In the same way, it is said to be in critical usage if all the stations are in critical usage.

A state of the network is a vector of dimension JM

$$\boldsymbol{n} = (n_{jm})_{1 \leq j \leq J, \ 1 \leq m \leq M}$$

where n_{jm} is the number of class j customers at station m.

The set of possible network states is then

$$\Omega = \left\{ \boldsymbol{n} \left| \sum_{m=1}^{M} n_{jm} = N_j \text{ for } j = 1, \cdots, J \right. \right\}.$$

Let $n_m = \sum_{j=1}^{J} n_{jm}$ and $\delta_m(n) = \begin{cases} 1 & \text{if } n \leq s_m \\ \prod_{i=1}^{n-s_m} \frac{s_m+i}{s_m} & \text{if } n > s_m \end{cases}$. Then the steady state probability for such a network is given by [Baskett et al. 1975]

$$\pi(\boldsymbol{n}) = \frac{1}{g} \delta(\boldsymbol{n})$$

with

$$\delta(\boldsymbol{n}) = \prod_{m=1}^{M} \delta_m(n_m) \prod_{j=1}^{J} \frac{\rho_{jm}^{n_{jm}}}{n_{jm}!},$$

where

$$g = \sum_{\boldsymbol{n} \in \Omega} \delta(\boldsymbol{n}) \tag{1}$$

is the normalization constant.

It is proven in [Ramakrishnan and Mitra 1982] that, in the case where each FCFS station has a unique server, that is $s_m = 1$ $(1 \leq m \leq L)$, this normalization constant g can be written as

$$g = \frac{1}{\prod_{j=1}^{J} N_j!} \int_{\boldsymbol{Q}^+} e^{-\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^{J} (\rho_{j0} + \boldsymbol{\rho}_j' \boldsymbol{u})^{N_j} d\boldsymbol{u}, \tag{2}$$

where

$$\begin{aligned} \boldsymbol{u} &= (u_1, \cdots, u_L)' \\ \mathbf{1} &= (1, \cdots, 1)' \\ \boldsymbol{\rho}_j &= (\rho_{j1}, \cdots, \rho_{jL})' \\ \boldsymbol{Q}^+ &= \{\boldsymbol{u} \in \mathbb{R}^L \ : \ u_l \geq 0, \ l = 1, \cdots, L\}. \end{aligned}$$

Given these normalization constants, network performance measures can be easily derived. For example, if g_j is the normalization constant of the network with one less class j customer,

$$TH_{jm} = \lambda_{jm} \frac{g_j}{g} \tag{3}$$

represents the throughput of class j customers at station m.

For complex networks, relation (1) can require a huge amount of computation in practice. In the same way, the integral form (2) is not computable in general. We then apply Monte-Carlo methods with importance sampling to evaluate them. The application of Monte Carlo summation and integration to the evaluation of the normalization constant was introduced in [Ross et al. 1994]. For general theory on Monte Carlo methods and importance sampling, see [Hammersley and Handscomb 1964]. We do not describe the technique called *integration with truncated normal sampling* (see [Ross and Wang 1997][Ross and Wang 1993]), because it is not applicable to our variance reduction algorithm.

2.2 Monte Carlo summation

2.2.1 *Summation with decomposition sampling method.* Define a probability p on Ω, and let $p(\boldsymbol{n})$ be the probability of state $\boldsymbol{n}$. We can write

$$g = \sum_{\boldsymbol{n} \in \Omega} \frac{\delta(\boldsymbol{n})}{p(\boldsymbol{n})} p(\boldsymbol{n}). \tag{4}$$

If $(\boldsymbol{n}^{(i)})_{1 \leq i \leq I}$ is a sample of mutually independent random variables $(\boldsymbol{N}^{(i)})_{1 \leq i \leq I}$ with probability distribution p, an estimator of g is

$$\bar{G}_{SD} = \frac{1}{I} \sum_{i=1}^{I} \frac{\delta(\boldsymbol{n}^{(i)})}{p(\boldsymbol{n}^{(i)})}$$

and, for I large enough, a confidence interval at risk approximately α is given by

$$\left[\bar{G}_{SD} - c_\alpha \frac{\sigma_p(\delta/p)}{\sqrt{I}}, \bar{G}_{SD} + c_\alpha \frac{\sigma_p(\delta/p)}{\sqrt{I}}\right],$$

where $\sigma_p^2(\delta/p)$ is the variance of random variable $\frac{\delta}{p}(\boldsymbol{N}^{(1)})$ under probability p (this variance is unknown but is easily estimated by its standard unbiased estimator) and $c_\alpha = \Phi^{-1}(1-\alpha/2)$ with Φ distribution function of the normal law with mean 0 and variance 1. It is argued in [Ross et al. 1994] and in [Ross and Wang 1993] that a good choice for the probability p is such that sampling is independent across classes,

$$p(\boldsymbol{n}) = \prod_{j=1}^{J} p_j(n_{j1}, \cdots, n_{jM}),$$

where

$$p_j(n_{j1}, \cdots, n_{jM}) = \frac{1}{K_j}\left(\prod_{m=1}^{L} \rho_{jm}^{n_{jm}}\right)\left(\prod_{m=L+1}^{M} \frac{\rho_{jm}^{n_{jm}}}{n_{jm}!}\right)$$

with $(n_{j1}, \cdots n_{jM})$ state vector for class j. Here the normalization constants K_j are easy to compute using convolution and recurrence (see [Ross et al. 1994]).

2.2.2 *Summation with rejection sampling method.* Summation with rejection is another importance sampling technique. Let

$$\Omega' = \{\boldsymbol{n} | n_{j1} + \cdots + n_{jL} \leq N_j, j = 1, \cdots J\}.$$

In [Ross and Wang 1997], Ross and Wang have modified (1) into

$$g = c \sum_{\boldsymbol{n}\in\Omega'} \left[\prod_{m=1}^{L} \frac{\delta_m(n_m)}{n_m!} \frac{n_m!}{n_{1m}! \cdots n_{Jm}!} \prod_{j=1}^{J} \left(\frac{N_j \rho_{jm}}{\rho_{j0}}\right)^{n_{jm}}\right] \prod_{j=1}^{J} \sigma(N_j, n_{j1} + \cdots + n_{jL}) \tag{5}$$

where $c = \prod_{j=1}^{J} \frac{\rho_{j0}^{N_j}}{N_j!}$ and $\sigma(N,n) = \prod_{i=N-n+1}^{N} \frac{i}{N}$. As in (4), we can use importance sampling for the computation of (5). It is proven in [Ross and Wang 1997] that a good choice for p is $p = p_\gamma$ where

$$p_\gamma(\boldsymbol{n}) = \prod_{m=1}^{L} \frac{\delta_m(n_m)\gamma_m^{n_m}}{n_m!\alpha_m} \frac{n_m!}{n_{1m}! \cdots n_{Jm}!} \left(\frac{\gamma_{1m}}{\gamma_m}\right)^{n_{1m}} \cdots \left(\frac{\gamma_{Jm}}{\gamma_m}\right)^{n_{Jm}}, \ \boldsymbol{n} \in \Lambda,$$

with

$$\begin{aligned} \Lambda &= \mathbb{N}^{JL} \text{ overset of } \Omega', \\ \boldsymbol{\gamma} &= (\gamma_{jm})_{1\leq j\leq J,\ 1\leq m\leq L}, \\ \gamma_m &= \gamma_{1m} + \cdots + \gamma_{Jm}, \\ \alpha_m &= \sum_{n=0}^{T_m} \frac{\delta_m(n)}{n!}\gamma_m^n, \end{aligned}$$

where T_m is the maximum of customers that can be present at station m. The best asymptotic choice for $\boldsymbol{\gamma}$ is described in [Ross and Wang 1997]. Then the i^{th} realization of g is

$$Z^{(i)} = c\alpha_1 \cdots \alpha_L 1_{\Omega'}(\boldsymbol{n}^{(i)}) \left[\prod_{m=1}^{L} \prod_{j=1}^{J} \left(\frac{N_j \rho_{jm}}{\gamma_{jm} \rho_{j0}} \right)^{n_{jm}^{(i)}} \right] \prod_{j=1}^{J} \sigma(N_j, n_{j1}^{(i)} + \cdots + n_{jL}^{(i)}).$$

2.3 Monte Carlo integration with exponential sampling

Let $(\boldsymbol{V}^i)_{1 \le i \le I}$ be a sequence of I mutually independent random vectors with probability density p defined on $\boldsymbol{Q}^+$. Let

$$Z^i = \frac{1}{\prod_{j=1}^{J} N_j!} \frac{e^{-\mathbf{1}'\boldsymbol{V}^i} \prod_{j=1}^{J} (\rho_{j0} + \boldsymbol{\rho}_j' \boldsymbol{V}^i)^{N_j}}{p(\boldsymbol{V}^i)}.$$

An estimator of g is $\bar{Z}^I = \frac{1}{I} \sum_{i=1}^{I} Z^i$ and a confidence interval can also be obtained. Ross, Tsang and Wang have proved (see [Ross et al. 1994]) that, asymptotically, the optimal importance sampling probability p is a product of exponential laws with parameters γ_l for $1 \le l \le L$. The value of γ_l depends on some network properties (see [Ross and Wang 1997]):

—if the network is in normal usage, then

$$\gamma_l = 1 - \sum_{j=1}^{J} N_j \frac{\rho_{jl}}{\rho_{j0}}, \ 1 \le l \le L;$$

—if the network is near or in critical usage, the values of the different γ_l will be estimated by means of a short preliminary simulation: having the estimated utilization, $util_l$, of the l^{th} FCFS station, we set

$$\gamma_l = 1 - util_l, \ 1 \le l \le L.$$

3. ANTITHETIC VARIATES

A simple description of this well known method can be found in [Hammersley and Handscomb 1964]. We apply it to the Monte Carlo integration with exponential sampling.

3.1 Application

Consider vectors $\boldsymbol{V}^{(1,b)}$ defined as the $\boldsymbol{V}^i$ in the section 2.3 and vectors $\boldsymbol{V}^{(2,b)}$ whose l^{th} coordinate is defined by $V_l^{(2,b)} = -\frac{\log(1 - e^{-\gamma_l V_l^{(1,b)}})}{\gamma_l}$. It is easy to show that the r.v. $\boldsymbol{V}^{(2,b)}$ is also exponentially distributed with parameter γ. Let

$$Z^{(1,b)} = f(\boldsymbol{V}^{(1,b)}) = \frac{1}{\prod_{j=1}^{J} N_j!} \frac{e^{-\mathbf{1}'\boldsymbol{V}^{(1,b)}} \prod_{j=1}^{J} (\rho_{j0} + \boldsymbol{\rho}_j' \boldsymbol{V}^{(1,b)})^{N_j}}{p(\boldsymbol{V}^{(1,b)})}$$

and

$$Z^{(2,b)} = f(\boldsymbol{V}^{(2,b)}) = \frac{1}{\prod_{j=1}^{J} N_j!} \frac{e^{-\mathbf{1}'\boldsymbol{V}^{(2,b)}} \prod_{j=1}^{J} (\rho_{j0} + \boldsymbol{\rho}_j' \boldsymbol{V}^{(2,b)})^{N_j}}{p(\boldsymbol{V}^{(2,b)})},$$

with $p(\boldsymbol{v}) = \prod_{l=1}^{L} (\gamma_l e^{-\gamma_l v_l})$ and $\gamma_l = 1 - \sum_{j=1}^{J} N_j \rho_{jl}/\rho_{j0}$. As the $\boldsymbol{V}^{(2,b)}$ have the same distribution as the $\boldsymbol{V}^{(1,b)}$, a new normalization constant estimator is

$$\bar{\mathcal{Z}}^B = \frac{1}{B} \sum_{b=1}^{B} \frac{1}{2} (Z^{(1,b)} + Z^{(2,b)}).$$

We have then

$$Var(\bar{\mathcal{Z}}^B) = Var(\bar{Z}^{2B}) + \frac{1}{2B} Cov(Z^{(1,1)}, Z^{(2,1)}),$$

where $\bar{Z}^{2B}$ is the standard estimator defined in 2.3, with $I = 2B$. This new estimator will be more efficient than the standard one if random variables $Z^{(1,b)}$ and $Z^{(2,b)}$ are negatively correlated.

THEOREM 1. *If the network is in normal usage, then the random variables $Z^{(1,b)}$ and $Z^{(2,b)}$ are negatively correlated.*

PROOF. To simplify the notations of the proof, let

$$\begin{aligned}
f(\boldsymbol{v}) = f(v_1, \cdots, v_L) &= e^{-(\mathbf{1}-\boldsymbol{\gamma})'\boldsymbol{v}} \prod_{j=1}^{J} (\rho_{j0} + \boldsymbol{\rho}_j' v)^{N_j} \\
f(\boldsymbol{v}^{(1)}) = f(v_1^{(1)}, \cdots, v_L^{(1)}) &= f\left(-\frac{\log(1 - e^{-\gamma_1 v_1})}{\gamma_1}, \cdots, -\frac{\log(1 - e^{-\gamma_L v_L})}{\gamma_L} \right) \\
a(\boldsymbol{v}) = a(v_1, \cdots, v_L) &= \prod_{l=1}^{L} \left(1 - e^{-\gamma_l v_l}\right)^{\frac{1-\gamma_l}{\gamma_l}} \\
b(\boldsymbol{v}) = b(v_1, \cdots, v_L) &= \prod_{j=1}^{J} \left(\rho_{j0} + \sum_{l=1}^{L} \rho_{jl} v_l \right)^{N_j} \\
c(\boldsymbol{v}) = c(v_1, \cdots, v_L) &= \prod_{j=1}^{J} \left(\rho_{j0} - \sum_{l=1}^{L} \rho_{jl} \frac{\log(1 - e^{-\gamma_l v_l})}{\gamma_l} \right)^{N_j}.
\end{aligned}$$

We have to show that

$$\int_{\boldsymbol{Q}^+} f(\boldsymbol{v}^{(1)}) f(\boldsymbol{v}) p(\boldsymbol{v}) d\boldsymbol{v} - \left(\int_{\boldsymbol{Q}^+} f(\boldsymbol{v}) p(\boldsymbol{v}) d\boldsymbol{v} \right)^2 \leq 0,$$

i.e.

$$\int_{\boldsymbol{Q}^+} \left(\prod_{l=1}^{L} e^{-v_l} \right) a(\boldsymbol{v}) b(\boldsymbol{v}) c(\boldsymbol{v}) d\boldsymbol{v} \leq \left(\int_{\boldsymbol{Q}^+} \left(\prod_{l=1}^{L} e^{-v_l} \right) b(\boldsymbol{v}) d\boldsymbol{v} \right)^2 = C^2,$$

with $C = \int_{\boldsymbol{Q}^+} \left(\prod_{l=1}^{L} e^{-v_l}\right) b(\boldsymbol{v}) d\boldsymbol{v}$. This is equivalent to showing

$$\int_{\boldsymbol{Q}^+} \left(\prod_{l=1}^{L} e^{-v_l}\right) b(\boldsymbol{v}) \left(C - a(\boldsymbol{v})c(\boldsymbol{v})\right) d\boldsymbol{v} \geq 0.$$

Given that $\left(\prod_{l=1}^{L} e^{-v_l}\right) b(\boldsymbol{v}) > 0$ for all $v_l > 0$ and all $1 \leq l \leq L$, it suffices to check that

$$a(\boldsymbol{v})c(\boldsymbol{v}) < C$$

for all $v_l > 0$ and $1 \leq l \leq L$. In that case, the integral will be positive, because we will integrate a positive function. Let

$$g(v_1, \cdots, v_L) = a(\boldsymbol{v})c(\boldsymbol{v}).$$

We are going to see that, under certain conditions, the maximum of g on $\boldsymbol{Q}^+$ is smaller than the value C. For all p we have

$$\frac{\partial g}{\partial v_p}(v_1, \cdots, v_L) = \frac{e^{-\gamma_p v_p}}{1 - e^{-\gamma_p v_p}} a(\boldsymbol{v})c(\boldsymbol{v}) \left[1 - \gamma_p - \sum_{k=1}^{J} \frac{N_k \rho_{kp}}{\rho_{k0} - \sum_{l=1}^{L} \rho_{kl} \frac{ln(1-e^{-\gamma_l v_l})}{\gamma_l}}\right]$$

The derivate has the same sign as the bracketed term. But this expression is positive: in normal usage, $\sum_{k=1}^{J} N_k \rho_{kp}/\rho_{k0} < 1 \ \forall p$, we set $\gamma_p = 1 - \sum_{k=1}^{J} N_k \rho_{kp}/\rho_{k0}$. Then, replacing γ_p by its value in the bracketed term, we obtain

$$\left[(1 - \gamma_p) - \sum_{k=1}^{J} \frac{N_k \rho_{kp}}{\rho_{k0} - \sum_{l=1}^{L} \rho_{kl} \frac{ln(1-e^{-\gamma_l v_l})}{\gamma_l}}\right]$$
$$= \left[\sum_{k=1}^{J} N_k \frac{\rho_{kp}}{\rho_{k0}} - \sum_{k=1}^{J} \frac{N_k \rho_{kp}}{\rho_{k0} - \sum_{l=1}^{L} \rho_{kl} \frac{ln(1-e^{-\gamma_l v_l})}{\gamma_l}}\right] > 0.$$

Thus we have

$$\forall 1 \leq l \leq L \ \forall (v_1, \cdots, v_L), \ \frac{\partial g}{\partial v_l}(v_1, \cdots, v_L) > 0.$$

As g is increasing, its maximum is obtained when $v_l \to +\infty, \ \forall 1 \leq l \leq L$. The upper bound of g is then

$$\lim_{\forall 1 \leq l \leq L \ v_l \to +\infty} g(v_1, \cdots, v_L) = \prod_{j=1}^{J} \rho_{j0}^{N_j}.$$

But

$$\begin{aligned} C &= \int_{\boldsymbol{Q}^+} \prod_{l=1}^{L} \left(e^{-v_l}\right) \prod_{j=1}^{J} \left(\rho_{j0} + \sum_{l=1}^{L} \rho_{jl} v_l\right)^{N_j} d\boldsymbol{v} \\ &> \int_{\boldsymbol{Q}^+} \prod_{l=1}^{L} \left(e^{-v_l}\right) \prod_{j=1}^{J} \rho_{j0}^{N_j} d\boldsymbol{v} \\ &= \prod_{j=1}^{J} \rho_{j0}^{N_j}. \end{aligned}$$

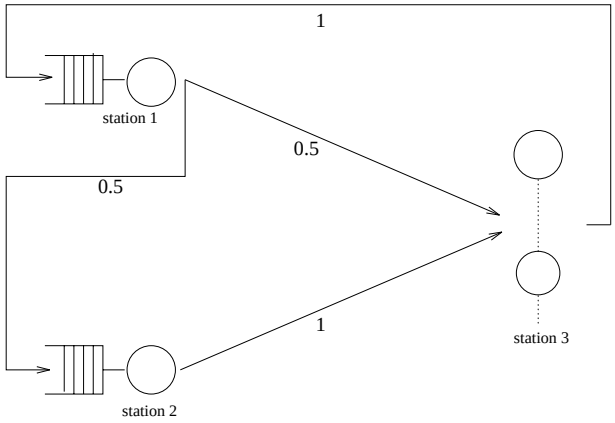


Fig. 1. Class 1

The property is then shown. □

Applying the method with antithetic variates in normal usage, we obtain a variance reduction. In the same way, we have investigated if the same approach deals to a variance reduction in the case of an utilization in critical usage, that is, if $\sum_{k=1}^{J} N_k \rho_{kp}/\rho_{k0} \geq 1$. We will see in an example that in this case, the property is not true in general: there can be an increase and not a reduction of the variance for some networks.

3.2 Numerical examples

The purpose here is to illustrate the variance reduction obtained with the new algorithm. We use small models to allow the reader to check the estimation values.

3.2.1 *Network in normal usage.* For the network that we study here, we use the following heuristic for all l:

$$\gamma_l = 1 - \sum_{k=1}^{J} N_k \frac{\rho_{kl}}{\rho_{k0}}.$$

The system consists of three stations: two FCFS and one IS. We consider two customer classes with five customers for class 1 and four customers for class 2.

The routing for customers of class 1 is described in Figure 1 and for customers of class 2 in Figure 2. The service rates are as follows:

$$\begin{aligned} \mu_1 &= 9.0, \\ \mu_2 &= 7.0, \\ \mu_{13} &= 0.5, \\ \mu_{23} &= 0.4. \end{aligned}$$

we have then $\sum_{k=1}^{2} N_k \rho_{k1}/\rho_{k0} = 181/450$ and $\sum_{k=1}^{2} N_k \rho_{k2}/\rho_{k0} = 57/140$.

We compare the confidence interval width obtained by Ross et al. for 10^6 iterations with those obtained for 5×10^5 iterations in the antithetic variates method, in order to have the same number of points of function f. As the computational time

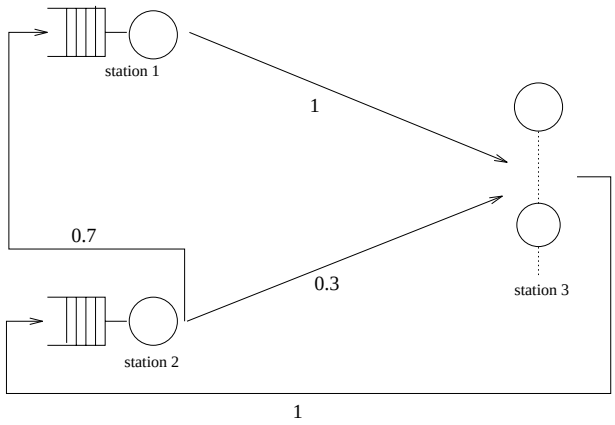


Fig. 2. Class 2

Table I. Confidence interval widths in normal usage. The estimated values are approximately TH_{11}=2.21, TH_{12}=1.10, TH_{21}=0.994 and TH_{22}=1.420.

Variate	Width for Ross *et al.*	Width with antithetic
TH_{11}	0.000634	0.000263
TH_{12}	0.000317	0.000132
TH_{21}	0.000282	0.000114
TH_{22}	0.000403	0.000163

is the same for both algorithms, the improvement will be manifest in the confidence interval reduction.

In Table I, the confidence interval width of the throughput of class j at station l TH_{lj} for $l, j = 1, 2$, are given.

As we can see, confidence interval widths are diminished by approximatively 2.5. It takes about $(2.5)^2 = 6.25$ more time using Ross et al. algorithm to obtain an interval with the same width.

3.2.2 *Network in critical usage.* Let us illustrate here that in this case, there can be an increase in the variance of the new estimator. The network presented here is in critical usage and needs a preliminary simulation for the estimation of the parameters γ_l in the software package MonteQueue 2.0. The network is still constituted of three stations, two FCFS with a unique server and one IS. On the other hand, we consider a single class of ten customers. Figure 3 gives the network routings. The service rates are as follows:

$$\begin{aligned} \mu_1 &= 0.5, \\ \mu_2 &= 0.5, \\ \mu_{13} &= 1.0. \end{aligned}$$

We have then $N_1\rho_{11}/\rho_{10} = N_1\rho_{12}/\rho_{10} = 20$.

We observe then in Table II an increase of the confidence interval width in the case of critical usage.

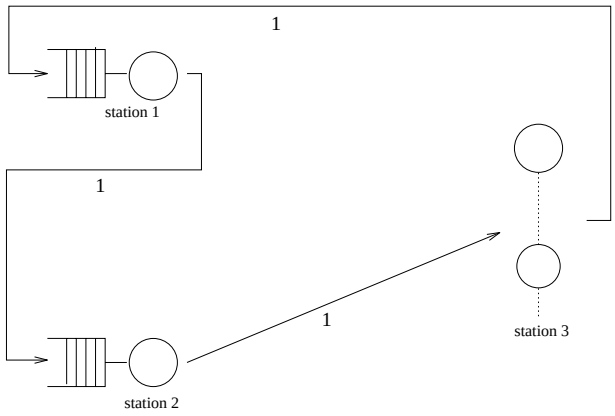


Fig. 3. Network in critical usage

Table II. Confidence Interval Widths in critical usage. The estimated values are approximately TH_{11}=0.542, TH_{12}=0.452.

Variate	Width for Ross *et al.*	Width with antithetic
TH_{11}	0.000629	0.000680
TH_{12}	0.000629	0.000680

4. QUASI-MONTE CARLO METHODS AND VARIANCE REDUCTION METHOD

4.1 Description

An alternative to Monte Carlo methods are Quasi-Monte Carlo ones. In the latters, we approximate $\int_{[0,1]^s} f(u)du$ by $\frac{1}{N}\sum_{n=1}^{N} f(\xi^{(n)})$ where $\mathcal{P} = (\xi^{(n)})_{n\epsilon\mathbb{N}}$ is a deterministic sequence. As a consequence, the error is deterministic as well. A measure of uniform distribution, which is necessary for convergence, is the *discrepancy*. Let $A_N(B,\mathcal{P})$ be the number of elements of $\mathcal{P}$ belonging to B among the N first elements of the sequence, that is $A_N(B,\mathcal{P}) = \sum_{n=1}^{N} 1_B(\xi^{(n)})$, and λ_s be the s-dimensional Lebesgue measure. Let $\mathcal{B}$ be the family of sets of form $\prod_{i=1}^{s}[0,u_i)$ with $u = (u_1,\cdots,u_s) \in [0,1)^s$. The discrepancy of the N first terms of $\mathcal{P}$ is defined by

$$D_N^*(\mathcal{P}) = \sup_{B\epsilon\mathcal{B}} \left| \frac{A_N(B,\mathcal{P})}{N} - \lambda_s(B) \right|.$$

Then $\mathcal{P}$ is uniformly distributed on $[0,1)^s$ if and only if $\lim_{N\to+\infty} D_N^*(\mathcal{P}) = 0$ [Kuipers and Niederreiter 1974].

For an infinite sequence $\mathcal{P}$, we can not have faster than $D_N^*(\mathcal{P}) = O(N^{-1}\log N)$ for $s = 1$, and $D_N^*(\mathcal{P}) = O(N^{-1}(\log N)^{\alpha(s)})$ for any s (see [Niederreiter 1992, pages 23-25], [Borel et al. 1991, page 10] or [Beck and Chen 1987] for a proof). The coefficient $\alpha(s)$ verifies $s/2 \le \alpha(s) \le s$ (and it is conjectured that $\alpha(s) = s$). Sequences with such convergence rate are called *low discrepancy sequences*.

One of the most known error bound is the following, called Koksma-Hlawka

Theorem [Zaremba 1968]:

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(\xi^{(n)}) - \int_{[0,1]^s} f(u)du \right| \leq V(f) D_N^*(\xi^{(1)}, \cdots, \xi^{(N)}), \tag{6}$$

where $V(f)$ is the variation in sense of Hardy and Krause [Niederreiter 1992; Zaremba 1968]. Asymptotically, convergence is then quicker than for standard Monte Carlo method which is in $O(1/\sqrt{N})$.

Unfortunately, the error bound of Koksma-Hlawka Theorem, given by equation (6), is difficult to evaluate in practice. As a matter of fact $V(f)$ is hard to compute or to over-bound, and there exists many functions with infinite variation for which Quasi-Monte Carlo convergence is fast [Bouleau and Lépingle 1993]. Furthermore, the cost of an exact computation of $D_N^*(\mathcal{P})$ [Niederreiter 1972] becomes quickly high and the computation is not feasible even for small values of the dimension and of the number of iterations. In the same way, an over-bound of type $C(\log N)^s/N$ is useless in a large dimension. For example in dimension $s = 10$, we need more than $N = 3 \times 10^{15}$ to have $(\log N)^s/N < 1$. Then, to obtain a useful error bound, we use low discrepancy sequences to reduce the variance in Monte Carlo methods.

First we are going to explain how we make a variance reduction for uniformly distributed random variable on set $[0,1)^s$, for a generic dimension s. Let X be a random variable uniformly distributed on $[0,1)^s$, and $(\xi^{(k)})_{k\in\mathbb{N}}$ a low discrepancy sequence. Let us consider the random variable

$$Z = \frac{1}{n} \sum_{k=1}^{n} f(\{X + \xi^{(k)}\}) \tag{7}$$

instead of $f(X)$, where $\{x\}$ is the fractional part for each coordinate of $x \in \mathbb{R}^s$. To compute f the same number of times, we have to compare the variance after I iterations with the variance after nI iterations in standard simulation. We will obtain a variance reduction if and only if

$$\sigma^2(\frac{1}{n} \sum_{k=1}^{n} f(\{X + \xi^{(k)}\})) < \frac{1}{n} \sigma^2(f(X)). \tag{8}$$

The first attempt in this direction can be found in [Cranley and Patterson 1976] where $(\xi^{(k)})_{k\in\mathbb{N}}$ is a lattice developed by Korobov and the second for Bayesian integration in [Shaw 1988] with low discrepancy sequences. Owen [Owen 1995][Owen 1994] uses a different technique, where the randomness is introduced on permutations, for Niederreiter sequences. In a recent paper [Tuffin 1996], we have compared Faure, SQRT, Niederreiter, Sobol and Halton sequences for a use of the random variable Z given in (7) on different classes of functions, and we have concluded that, generally, Sobol sequences [Sobol' 1967][Sobol' 1976] using the Gray code [Antonov and Saleev 1979] are the most efficient (in terms of estimated variance multiplied by the observed computational time).

To get an idea of the convergence speed, consider first the case of functions with bounded variation.

THEOREM 2. *Let $\mathcal{P} = (\xi^{(k)})_{k\in\mathbb{N}}$ be a low discrepancy sequence over $[0,1)^s$. If f*

is a function with bounded variation, we have

$$\sigma^2\left(\frac{1}{n}\sum_{k=1}^{n} f(\{X+\xi^{(k)}\})\right) = O(n^{-2}(\log n)^{2s}).$$

PROOF. See first that $E\left(\frac{1}{n}\sum_{k=1}^{n} f(\{\xi^{(k)}+X\})\right) = E(f(X))$. Next, let us define the sequence $\mathcal{Q}_x = (\{\xi^{(k)}+x\})_{k\in\mathrm{IN}}$. Note that, $\forall 1 \le i \le s$,

$$1_{[0,z_i)}(\{\xi_i^{(k)}+x_i\}) = \begin{cases} 1_{[0,z_i)}(\xi_i^{(k)}) & \text{if} \quad x_i = 0 \\ 1_{[1-x_i,1+z_i-x_i)}(\xi_i^{(k)}) & \text{if} \quad z_i - x_i \le 0 \text{ and } x_i > 0 \\ 1_{[1-x_i,1)}(\xi_i^{(k)}) + 1_{[0,z_i-x_i)}(\xi_i^{(k)}) & \text{if} \quad z_i > x_i > 0 \end{cases} \tag{9}$$

and similarly

$$\lambda_1([0,z_i)) = \begin{cases} \lambda_1([0,z_i)) & \text{if} \quad x_i = 0 \\ \lambda_1([1-x_i,1+z_i-x_i)) & \text{if} \quad z_i - x_i \le 0 \text{ and } x_i > 0 \\ \lambda_1([1-x_i,1)) + \lambda_1([0,z_i-x_i)) & \text{if} \quad z_i > x_i > 0. \end{cases}$$

Expanding then the products $1_{[0,z)}(\{\xi^{(k)}+x\}) = \prod_{i=1}^{s} 1_{[0,z_i)}(\{\xi_i^{(k)}+x_i\})$ and $\lambda_s([0,z)) = \prod_{i=1}^{s}\lambda_1([0,z_i))$, we obtain at most 2^s terms, so

$$\begin{aligned} D_n^*(\mathcal{Q}_x) &\le 2^s \sup_{u,v\in[0,1)^s,\ u<v} \left|\frac{A_n([u,v),\mathcal{P})}{n} - \lambda_s([u,v))\right| \\ &\le 4^s D_n^*(\mathcal{P}) \text{ [Niederreiter 1992, page 15].} \end{aligned} \tag{10}$$

Then,

$$\begin{aligned} \sigma^2\left(\frac{1}{n}\sum_{k=1}^{n} f(\{\xi^{(k)}+X\})\right) &= \int_{[0,1)^s}\left|\frac{1}{n}\sum_{k=1}^{n} f(\{x+\xi^{(k)}\}) - E(f)\right|^2 dx \\ &\le \int_{[0,1)^s} (V(f)D_n^*(\mathcal{Q}_x))^2 dx \\ &\le (4^s V(f) D_n^*(\mathcal{P}))^2, \end{aligned}$$

so we get the Theorem. □

The estimator in (7) gives also good results for functions with infinite variation, as in the case of quasi-Monte Carlo methods where [Wozniakowski 1991] the mean square error

$$\int_{\mathcal{F}}\left(\frac{1}{n}\sum_{k=1}^{n} f(\xi^{(k)}) - \int_{[0,1)^s} f(u)du\right)^2 d\mu_W(f)$$

obtained on the set $\mathcal{F}$ of continuous functions (equipped with the Wiener measure μ_W concentrated on functions with infinite variation [Morokoff and Caflisch 1994]) is equal to the square of the mean square discrepancy

$$T_n^{(2)}(\mathcal{Q}) = \left(\int_{[0,1)^s}\left(\frac{A_n([0,u),\mathcal{Q})}{n} - \lambda_s([0,u))\right)^2 du\right)^{1/2}$$

of $\mathcal{Q} = (1 - \xi^{(k)})_{k \in \mathbb{N}}$. Recall that the Wiener measure μ_W is Gaussian with mean value 0 and covariance kernel $R(x,y) = \int_{\mathcal{F}} f(x)f(y)d\mu_W(f) = \prod_{i=1}^{s} \min(x_i, y_i)$. The following result gives another view of the asymptotic convergence speed of this approach.

THEOREM 3. *For any low discrepancy sequence $\mathcal{P} = (\xi^{(k)})_{k \in \mathbb{N}}$, the mean variance $\sigma^2_{n,avg}$ of the estimator $\frac{1}{n}\sum_{k=1}^{n} f(\{\xi^{(k)} + X\})$, taken over the set $\mathcal{F}$ of continuous functions f on $[0,1]^s$ equipped with the Wiener measure μ_W, is in $O(n^{-2}(\log n)^{2s})$.*

PROOF.

$$\begin{aligned}
\sigma^2_{n,avg} &= \int_{\mathcal{F}} \sigma^2 \left(\frac{1}{n} \sum_{k=1}^{n} f(\{\xi^{(k)} + X\}) \right) d\mu_W(f) \\
&= \int_{[0,1)^s} \int_{\mathcal{F}} \left(\frac{1}{n} \sum_{k=1}^{n} f(\{\xi^{(k)} + x\}) - \int_{[0,1)^s} f(u)du \right)^2 d\mu_W(f)dx \\
&= \int_{[0,1)^s} (T_n^{(2)}((1 - \{x + \xi^{(k)}\})_{k \in \mathbb{N}}))^2 dx
\end{aligned}$$

by Wozniakowski's result. Let $D_n(\mathcal{I}, \mathcal{P}) = \sup_{B \epsilon \mathcal{I}} |A_n(B, \mathcal{P})/n - \lambda_s(B)|$, with $\mathcal{I}$ set of subintervals of $[0,1]^s$ such that both open, closed or mixed intervals are allowed. We get $\forall y \in [0,1)^s$,

$$\begin{aligned}
\left| A_n([0,y), (1 - \{x + \xi^{(n)}\}))/n - \prod_{i=1}^{s} y_i \right| &= \left| A_n((1-y,1], (\{x + \xi^{(n)}\}))/n - \prod_{i=1}^{s} y_i \right| \\
&\leq 2^s D_n(\mathcal{I}, \mathcal{P})
\end{aligned}$$

as $1_{(1-y_i,1]}(\{\xi_i^{(k)} + x_i\}) = 1_{(1-y_i,1)}(\{\xi_i^{(k)} + x_i\})$ and using the same kind of arguments as in (10), but with

$$1_{(1-y_i,1)}(\{\xi_i^{(k)} + x_i\}) = \begin{cases} 1_{(1-y_i-x_i,1-x_i)}(\xi_i^{(k)}) & \text{if } 1 - y_i - x_i \geq 0 \\ 1_{(2-y_i-x_i,1)}(\xi_i^{(k)}) + 1_{[0,1-x_i)}(\xi_i^{(k)}) & \text{otherwise} \end{cases}$$

instead of (9). This implies that $(T_n^{(2)}((1 - \{x + \xi^{(k)}\})_{k \in \mathbb{N}}))^2 \leq (2^s D_n(\mathcal{I}, \mathcal{P}))^2$. But $D_n(\mathcal{I}, \mathcal{P})$ is also in $O(n^{-1}(\log n)^s)$ (using Koksma-Hlawka theorem and taking 2^s as an upper bound of the variation of the indicator function of each subinterval of $[0,1]^s$). □

The complete study of the convergence speed of the approach remains to be done. In some recent work [Owen 1996], it has been shown that under some conditions on the integrand, the variance of Owen's method is $O(n^{-3}(\log n)^{s-1})$. It is shown numerically in [Tuffin 1996] that for different functions the method using relation (7) performs generally better.

In our applications, we do not use integration over uniform law on $[0,1)^s$, but over more general laws. We can remark that, by a simple transformation, it is often possible to generate a sequence with a particular distribution on a subset of $\mathbb{R}^s$ from a uniformly distributed sequence on $[0,1)^s$ as we have just made, by means of s one-dimensional pseudo-inverse functions, as done by Ross *et al.* For example, for

integration with exponential sampling (see sub-section 2.3), the L coordinates are independent and we can then generate them separately. Thus the mathematical dimension is $s = L$. If $F_l^{-1}(u) = -\log(1-u)/\gamma_l$ is the pseudo-inverse function of the distribution of an exponential law with parameter γ_l $(l = 1, \cdots, L)$, a new estimator of g is

$$\bar{Z}^I = \frac{1}{I}\sum_{i=1}^{I} Z^i.$$

with

$$Z^i = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{\prod_{j=1}^{J} N_j!} \frac{e^{-\mathbf{1}'\boldsymbol{V}^{(i,k)}} \prod_{j=1}^{J} (\rho_{j0} + \boldsymbol{\rho}_j' \boldsymbol{V}^{(i,k)})^{N_j}}{p(\boldsymbol{V}^{(i,k)})}$$

and

$$\boldsymbol{V}^{(i,k)} = (F_1^{-1}(\{U_1^{(i)} + \xi_1^{(k)}\}), \cdots, F_L^{-1}(\{U_L^{(i)} + \xi_L^{(k)}\}))$$

for $(U_l^{(i)})_{1 \le l \le L, 1 \le i \le I}$ independent uniform random variable on $[0,1)$.

For summation with decomposition sampling, sampling over classes is independent, and sampling for class j in station m $(1 \le m \le M)$ is found conditionally to the variables $(n_{jl})_{1 \le l \le m-1}$ (see [Ross et al. 1994]). Then p, defined as in 2.2.1, is generated from a uniform random variable on $[0,1)^{JM}$. Let $U_{jm}^{(i)}$ with $1 \le l \le M$, $1 \le j \le J$ and $1 \le i \le I$ be independent uniform random variables on $[0,1)$. A new estimator of g is

$$\bar{G}_{SD} = \frac{1}{I}\sum_{i=1}^{I} \left(\frac{1}{n}\sum_{k=1}^{n} \frac{\delta(\boldsymbol{n}^{(i,k)})}{p(\boldsymbol{n}^{(i,k)})} \right)$$

with $\boldsymbol{n}^{(i,k)}$ sampled (from $\{U^{(i)} + \xi^{(k)}\}$ instead of $U^{(i)}$) by means of conditional probabilities [Ross et al. 1994].

For summation with rejection sampling, the number V_m of customers in station m $m \le L$ is first selected (with the help of a pseudo-inverse function). This number must be less than T_m.Then each of the V_m customers is chosen to be of class j with probability γ_{jm}/γ_m. The mathematical dimension in this technique is $L + \sum_{m=1}^{L} T_m$. A new summation with rejection sampling estimator of g is

$$\bar{Z}^I = \frac{1}{I}\sum_{i=1}^{I} Z^i$$

with

$$Z^i = \frac{1}{n}\sum_{k=1}^{n} c\alpha_1 \cdots \alpha_L 1_{\Omega'}(\boldsymbol{n}^{(i,k)}) \left[\prod_{m=1}^{L} \prod_{j=1}^{J} \left(\frac{N_j \rho_{jm}}{\gamma_{jm} \rho_{j0}} \right)^{n_{jm}^{(i,k)}} \right] \prod_{j=1}^{J} \sigma\left(N_j, \sum_{l=1}^{L} n_{jl}^{(i,k)} \right).$$

$n_m^{(i,k)}$ is chosen from $\{U_m^{(i)} + \xi_m^{(k)}\}$ on the set $\{0, \cdots, T_m\}$ by means of the pseudo-inverse of the probability $\delta_m(n_m^{(i,k)})/\alpha_m(n_m^{(i,k)})!$. We obtain $n_{jm}^{(i,k)}$ by distributing from $\{U_r^{(i)} + \xi_r^{(k)}\}$ $(r > L)$ the $n_m^{(i,k)}$ customers among the J classes: a customer is of class j with probability γ_{jm}/γ_m.

Table III. Confidence interval widths for integration with exponential sampling. The estimated values are approximately g=9.940e+64, g_j=9.55e+62 and TH_{jm}=9.6095e-03.

Variate	Ross *et al.* $I = 10^6$	$n = 10^2$, $I = 10^4$	$n = 10^3$, $I = 10^3$	$n = 10^4$, $I = 10^2$
g	8.1970e+60	3.0134e+60	1.4013e+60	6.9722e+59
g_j	1.3187e+59	4.3284e+58	1.8322e+58	8.1147e+57
TH_{jm}	5.4623e-07	1.6259e-07	6.4567e-08	2.5715e-08

Table IV. Confidence interval widths for summation with decomposition sampling. The estimated values are approximately g=1.72e+15, g_1=3.18e+14 and TH_{1m}=1.849e-01.

Variate	Ross *et al.* $I = 10^6$	$n = 10^2$, $I = 10^4$	$n = 10^3$, $I = 10^3$	$n = 10^4$, $I = 10^2$
g	2.7575e+12	2.3697e+12	2.2816e+12	2.2489e+12
g_1	4.6949e+11	4.2516e+11	4.3404e+11	3.7260e+11
TH_{1m}	1.1949e-03	1.0306e-03	9.3194e-04	7.0177e-04

4.2 Numerical results

We apply this technique to the estimation of normalization constants in the three methods described in previous section. We will compare all the methods with Ross *et al.* ones for different values of n.

In all our examples the routing is supposed to be the same for each class. Customers go from station m to station $m+1$ ($m = 1, \cdots, M-1$) and from station M to station 1 with probability 1. We give different values to μ_m ($1 \leq m \leq L$), J, $N_1, \cdots, N_J$, μ_{jm} ($L+1 \leq m \leq M$) and s_m ($1 \leq m \leq L$) in order to be in the case where each method of Ross *et al.* performs the best (see [Ross and Wang 1997]).

To test integration with exponential sampling method, we set $M = 9$, $L = 7$, $\mu_m = 1.0$ ($1 \leq m \leq L$), $J = 15$, $N_1 = \cdots = N_J = 2$, $\mu_{jm} = 0.01$ ($L+1 \leq m \leq M$) and $s_m = 1$ ($1 \leq m \leq L$). The results (confidence interval width) for MonteQueue for $I = 10^6$ iterations and for our algorithm for $I = 10^4$ and $n = 10^2$, for $I = 10^3$ and $n = 10^3$ and for $I = 10^2$ and $n = 10^4$, then with the same number of calls of the function, are given in Table III.

We can observe for integration with exponential sampling technique, a large variance reduction. For $n = 10^4$ and $I = 10^2$, we obtain a standard deviation reduction of about 21.25 for the estimation of the throughput. Then it requires about $(21.25)^2 \approx 451$ times more iterations to give an interval with the same width with Ross, Tsang and Wang's method. Moreover, the computational time is about 0.72 that of Ross, Tsang and Wang's method. This is not due to the smaller number of calls to the random generator, but to the fact that we need less operations to compute the variance and the covariance. As a matter of fact the latter variables are computed only I times instead of nI times as in a pure Monte Carlo technique. We have thus a gain on two points: variance and time. The efficiency $1/(\sigma^2 \times t)$ is then 626 times better for the new technique.

The results for summation with decomposition sampling for $M = 9$, $L = 7$, $J = 3$, $N_1 = N_2 = N_3 = 5$, $s_m = 3$, $\mu_m = 1.0$ ($1 \leq m \leq L$) and $\mu_{jm} = 0.1$ ($L+1 \leq m \leq M$) are in Table IV. We give the confidence interval width only for the first class because the results for the other classes are of the same order.

The improvements are smaller in this case (the interval width for TH_{1m} with

Table V. Confidence interval widths for summation with rejection sampling. The estimated values are approximately g=3.17e+24, g_1=7.39e+22 and TH_{1m}=2.325e-02.

Variate	Ross *et al.* $I = 10^6$	$n = 10^2$, $I = 10^4$	$n = 10^3$, $I = 10^3$	$n = 10^4$, $I = 10^2$
g	2.5257e+21	3.6708e+21	2.7548e+21	1.6920e+21
g_1	1.0017e+20	9.8506e+19	7.2692e+19	4.8122e+19
TH_{1m}	2.5527e-05	2.0094e-05	1.3921e-05	1.0267e-05

$n = 10^4$ and $I = 10^2$ if about 0.58 the initial interval width, then it requires about $(1/0.58)^2 = 2.97$ times more iterations for Ross, Tsang and Wang to perform the same interval width). As the computational time is 0.80 times that of Ross, Tsang and Wang's, the efficiency is 3.7 times better for the new algorithm. Nevertheless, we can observe that as n, the number of elements of the low discrepancy sequence, increases, the variance reduces.

The results for summation with rejection sampling for $M = 9$, L=3, $J = 15$, $N_j = 1$ $(1 \leq j \leq J)$, and $s_m = 3$, $\mu_m = 1.0$ $(1 \leq m \leq L)$ and $\mu_{jm} = 0.15$ $(L+1 \leq m \leq M)$, are in Table V. As an example, we give the confidence interval width of the throughput only for the first class, instead of enumerating it for all the fifteen classes.

The variance reduction is larger as the mathematical dimension is smaller. We can observe that the interval width for TH_{1m} with $n = 10^4$ and $I = 10^2$ is 0.40 times the interval width with Ross and Wang's method (which need about six times more iteration to obtain the same interval width). As the computational time is 0.75 times that of Ross and Wang's, the efficiency is 8.25 times better with the use of low discrepancy sequences. For a sufficiently large number n of calls of elements of the low discrepancy sequence, we get a variance reduction. Moreover, as we increase n (and so decrease I, the number of random variables), the variance reduction is more important. But we have to take care to keep a sufficiently large I to invoke the normal approximation.

The fact that variance reduction is larger for integration with exponential sampling is due to the smaller mathematical dimension. As a matter of fact, in our examples, dimension for integration with exponential sampling is $L = 7$, whereas it is $JM = 18$ for summation with decomposition and $L + \sum_{l=1}^{L} T_l = 48$ for summation with rejection. Observe that if $\sigma^2\left(n^{-1}\sum_{i=1}^{n} f(\{X + \xi^{(i)}\})\right) = O(n^{-2}(\log n)^{2s})$, given that $D_n^*((\xi^{(k)})_{k\in\mathbb{N}}) \geq C_s(\log n)^{s/2}/n$, it seems reasonable to conjecture that, generally, the convergence speed of the variance decreases exponentially with dimension s. Then, although the variance reduction is very large asymptotically, it requires a larger n as the dimension increases.

5. CONCLUSION

We describe two new applications of variance reduction methods to product form queuing networks: the first is based on antithetic variates and the second on low discrepancy sequences. These methods are easy to implement. Antithetic variates techniques always give an improvement over integration with exponential sampling in the case of normal usage, that is when the latter method performs the best. The technique based on low discrepancy sequences gives better results when the mathematical dimension is small. As the mathematical dimension is usually small

for integration with exponential sampling, we recommend its use in this case (our example gives an efficiency 626 times better than with Ross, Tsang and Wang's method). In the case of summation, either with decomposition or rejection sampling, the use of our methods depends on two factors. If the dimension is not very large and the needed interval width is small (then a large number of iterations is required), we can apply our methods with a large n. If the mathematical dimension is large and the number of iterations needed is small, we recommend the use of Ross *et al.*'s algorithms.

ACKNOWLEDGMENTS

The author would like to thank the editors and the anonymous referees for their valuable comments and suggestions.

REFERENCES

Antonov, I. and Saleev, V. 1979. An economic method of computing lp_τ-sequences. *USSR Computational Math. and Math. Phys. 19*, 1, 252–256.

Baskett, F., Chandy, M., Muntz, R., and Palacios, J. 1975. Open, closed and mixed networks of queues with different classes of customers. *Journal of the Association for Computing Machinery 22*, 248–260.

Beck, J. and Chen, W. 1987. *Irregularities of Distribution.* Cambridge University Press.

Borel, J., Pagès, G., and Xiao, Y. 1991. *Probabilités numériques*, Chapter suites à discrépance faible et intégration numérique. INRIA. collection didactique.

Bouleau, N. and Lépingle, D. 1993. *Numerical methods for stochastic processes.* John Wiley and Sons.

Buzen, J. 1973. Computational algorithms for closed queueing networks with exponential servers. *Communications of ACM 16*, 527–531.

Cranley, R. and Patterson, T. 1976. Randomization of number theoretic methods for multiple integration. *SIAM J. Numer. Anal. 13*, 6 (December), 904–914.

Hammersley, J. M. and Handscomb, D. C. 1964. *Monte Carlo Methods.* Methuen, London.

Kuipers, L. and Niederreiter, H. 1974. *Uniform Distribution of Sequences.* John Wiley, New York.

Mckenna, J. and Mitra, D. 1982. Integral Representations and Asymptotic Expansions for Closed Markovian Queueing Networks: Normal Usage. *Bell Systems Technical Journal 61*, 5 (May-June), 661–683.

McKenna, J. and Mitra, D. 1984. Asymptotic Expansions and Integral Representations of Moments of Queue Lenghts in Closed Markovian Networks. *Journal of the Association for Computing Machinery 31*, 2 (April), 346–360.

Mckenna, J., Mitra, D., and Ramakrishnan, K. G. 1981. A Class of Closed Markovian Queuing Networks: Integral Representations, Asymptotic Expansions, and Generalizations. *The Bell Systems Technical Journal 60*, 5 (May-June), 599–641.

Morokoff, W. J. and Caflisch, R. E. 1984. Quasi-Random Sequences and their Discrepancies. *SIAM Journal on Scientific Computing*, (December), 1571–1599.

Niederreiter, H. 1972. Discrepancy and Convex Programming. *Annali di Matematica 93*, 89–97.

Niederreiter, H. 1978. Quasi-Monte Carlo Methods and Pseudo-Random Numbers. *Bulletin of the American Mathematical Society 84*, 957–1041.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods.* CBMS-SIAM 63, Philadelphia.

Owen, A. B. 1994. Monte Carlo Variance of Scrambled Equidistribution Quadrature. Technical report, Department of Statistics, Stanford University.

Owen, A. B. 1995. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In H. Niederreiter and P. J.-S. Shiue Eds., *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Volume 106 of *Lecture Notes in Statistics* (1995), pp. 299–315. Springer.

Owen, A. B. 1996. Scrambled Net Variance for Integrals of Smooth Functions. Technical report, Department of Statistics, Stanford University.

Ramakrishnan, K. and Mitra, D. 1982. An Overview of PANACEA, a Software Package for Analyzing Markovian Queueing Networks. *Bell System Technical Journal 61*, 2849–2872.

Reiser, M. and Kobayashi, H. 1975. Queueing networks with multiple closed chains: Theory and computational algorithms. *IBM J. of Research and Development 19*, 283–294.

Ross, K., Tsang, D., and Wang, J. 1994. Monte Carlo summation and integration applied to multichain queueing networks. *Journal of the Association for Computing Machinery 41*, 6 (November), 1110–1135.

Ross, K. and Wang, J. 1993. Asymptotically Optimal Importance Sampling for Product-Form Queueing Networks. *ACM Transactions on Modeling and Computer Simulation 3*, 244–268.

Ross, K. and Wang, J. 1997. Implementation of Monte Carlo Integration for the Analysis of product-form queueing networks. *Performance Evaluation 29*, 4, 273–292.

Shaw, J. E. H. 1988. A quasirandom approach to integration in Bayesian statistics. *Ann. Statist. 16*, 895–914.

Sobol', I. 1967. The distribution of points in a cube and the approsimate evaluation of integrals. *USSR Computational Math. and Math. Phys. 7*, 86–112.

Sobol', I. 1976. Uniformly distributed sequences with an additional uniform property. *USSR Computational Math. and Math. Phys. 16*, 5, 236–242.

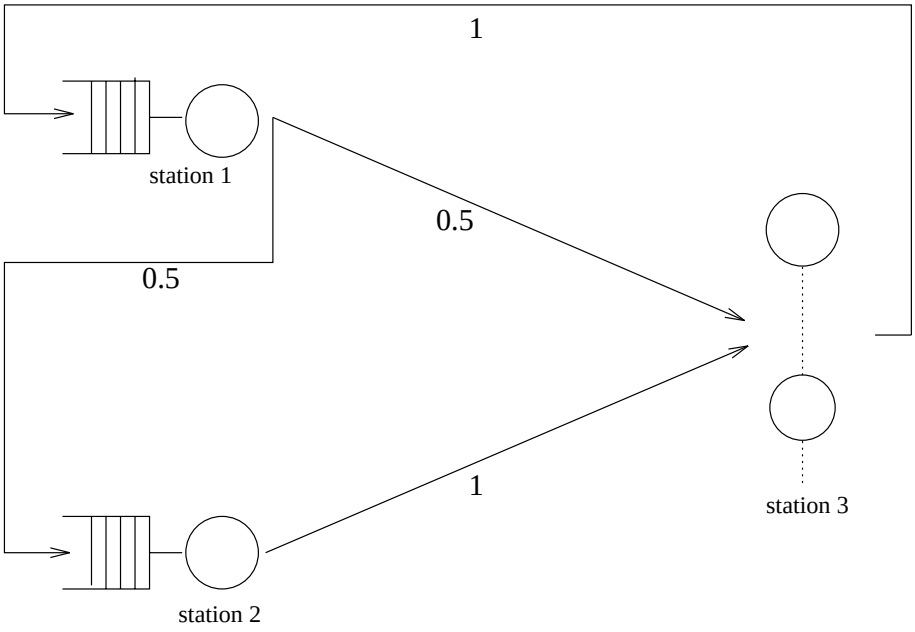
Tuffin, B. 1996. On the Use of Low Discrepancy Sequences in Monte Carlo Methods. *Monte Carlo Methods and Applications 2*, 4, 295–320.

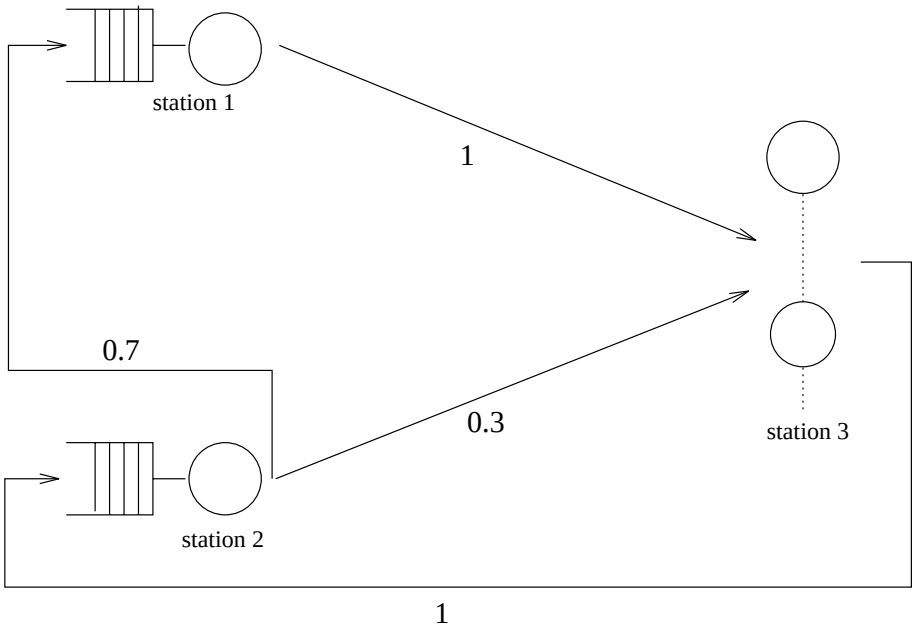
Wang, J. and Ross, K. 1994. Asymptotic Analysis for Closed Multiclass Queueing Networks in Critical Usage. *Queueing Systems: Theory and Applications 16*, 167–191.

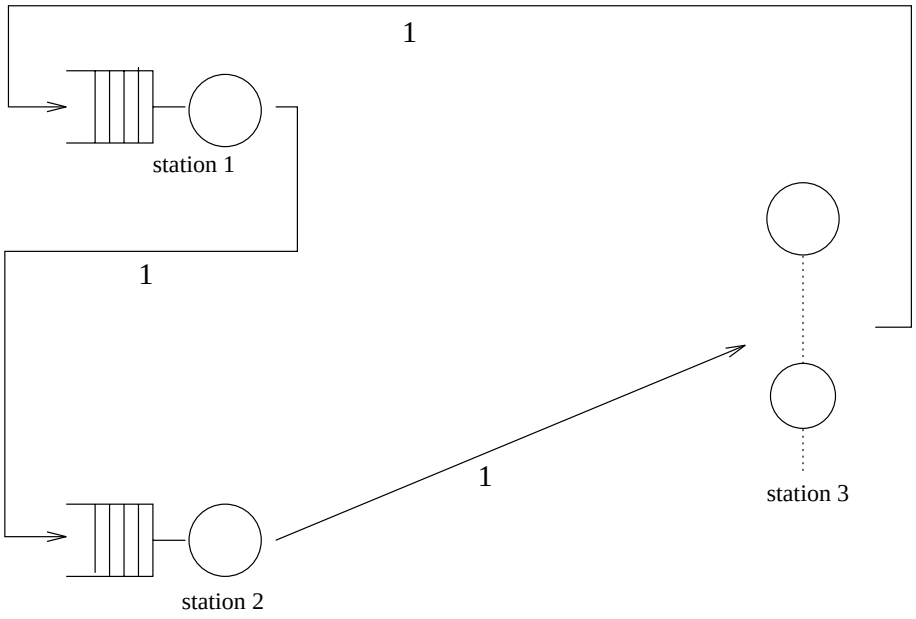
Wozniakowski, H. 1991. Average Case Complexity of Multivariate Integration. *Bulletin of the American Mathematical Society 24*, 1 (January), 185–194.

Zaremba, S. K. 1968. Some applications of multidimensionnal integration by parts. *Ann. Pol. Math XXI*, 85–96.

Variate	Width for Ross *et al*	Width with antithetic
TH_{11}	0.000634	0.000263
TH_{12}	0.000317	0.000132
TH_{21}	0.000282	0.000114
TH_{22}	0.000403	0.000163


1
station 1
0.5
0.5
1
station 3
station 2


station 1
1
0.7
0.3
station 3
station 2
1


1
station 1
1
1
station 3
station 2

Variate	Width for Ross *et al*	Width with antithetic
TH_{11}	0.000629	0.000680
TH_{12}	0.000629	0.000680

Variate	Ross *et al* $I = 10^6$	$n = 10^2$, $I = 10^4$	$n = 10^3$, $I = 10^3$	$n = 10^4$, $I = 10^2$
g	8.1970e+60	3.0134e+60	1.4013e+60	6.9722e+59
g_j	1.3187e+59	4.3284e+58	1.8322e+58	8.1147e+57
TH_{jm}	5.4623e-07	1.6259e-07	6.4567e-08	2.5715e-08

Variate	Ross *et al* $I = 10^6$	$n = 10^2$, $I = 10^4$	$n = 10^3$, $I = 10^3$	$n = 10^4$, $I = 10^2$
g	2.7575e+12	2.3697e+12	2.2816e+12	2.2489e+12
g_1	4.6949e+11	4.2516e+11	4.3404e+11	3.7260e+11
TH_{1m}	1.1949e-03	1.0306e-03	9.3194e-04	7.0177e-04

Variate	Ross *et al* $I = 10^6$	$n = 10^2$, $I = 10^4$	$n = 10^3$, $I = 10^3$	$n = 10^4$, $I = 10^2$
g	2.5257e+21	3.6708e+21	2.7548e+21	1.6920e+21
g_1	1.0017e+20	9.8506e+19	7.2692e+19	4.8122e+19
TH_{1m}	2.5527e-05	2.0094e-05	1.3921e-05	1.0267e-05