行政院國家科學委員會專題研究計畫 成果報告

網路模型的快速模擬方法研究
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 95-2221-E-004-003-
執 行 期 間 ：95 年 08 月 01 日至 96 年 07 月 31 日
執 行 單 位 ：國立政治大學資訊管理學系

計 畫 主 持 人：謝明華

計畫參與人員：博士班研究生-兼任助理：邱于芬、邱耀漢

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 96 年 12 月 17 日

行政院國家科學委員會專題研究計畫成果報告

網路模型的快速模擬方法研究

計畫類別： ☑ 個別型計畫　　☐ 整合型計畫

計畫編號： NSC 95-2221-E-004-003-

執行期間： 中華民國95年8月1日至96年7月31日

計畫主持人：謝明華

執行單位：國立政治大學資訊管理學系

中華民國96年11月30日

網路模型的快速模擬方法研究

計畫編號： NSC 95-2221-E-004-003-

執行期間： 中華民國95年8月1日至96年7月31日

主持人： 謝明華

E-mail: mhsieh@mis.nccu.edu.tw

執行單位： 國立政治大學資訊管理學系

一、中文摘要

對改進偶發事件 (rare event) 模擬的效率, 重點抽樣這種模擬技術通常是最好的選擇。過去, 在佇列 (queueing) 與可靠度 (reliability) 模型上, 應用重點抽樣成功的例子不少。但是, 對一特定的模型, 選取一個有效率的重點抽樣分配, 通常是困難與費時的。爲解決這樣的難題, 我們針對佇列模型開發出一種能自我調適 (adaptive) 的演算法來選取好的重點抽樣分配。我們進行的數值例子顯示我們提出的演算法確能有效的減少估計的變異數。

關鍵詞：自我調適的演算法, 重點抽樣, 偶發事件, 佇列網路, 模擬, 變異數縮減

Abstract

To improve the efficiency of rare event simulation, the technique of importance sampling is usually the best candidate. A number of successful examples of applying importance sampling to rare event simulation exist in queueing and reliability models. However, it is usually difficult and time consuming to select an effective importance sampling distribution for a particular model. Toalleviate such difficulty, we developed an adaptive algorithm for selecting effective importance sampling distributions for queueing models. The numerical examples we conducted showed that the porposed algorithm is effectiveness in reducing estimator's variance.

Keywords: Adaptive algorithm, importance sampling, rare event, queueing networks, simulation, variance reduction

二、研究目的與文獻探討

We develop an adaptive algorithm for estimating probabilities of rare, but significant, events that appear in performance evaluation of queueing models. In particular, we wish to develop an algorithm to estimate buffer overflow probabilities in a generalized open Jackson network - open networks of GI/G/1 queues with Markovian routing. The probability we consider will be π_K, the steady-state probability that the total population in the network exceed K. This type of problem has received considerable attention in the literature, see, e.g., (Parekh and Walrand, 1989; Glasserman and Kou, 1995; Frater et al., 1991; Frater and Anderson, 1994). It is very inefficient or sometimes impossible to estimate small probabilities via naive simulation. Hence, techniques for accelerating simulation speed are essential. Importance sampling (Hammersley and Handscomb, 1965; Glynn and Iglehart, 1989) is such a technique.

A number of successful examples of applying importance sampling to rare event simulation exist in queueing and reliability models (see Heidelberger (1995); Shahabuddin (1995); Asmussen and Rubinstein (1995) for surveys.) However, it is difficult and time consuming to select an effective importance sampling distribution for a particular model. To alleviate such difficulty, we propose an adaptive algorithm for selecting effective importance sampling distributions for queueing models. The algorithm is based on the conditioned limit theorem for random walks. It computes a good importance sampling distribution via the samples obtained under certain conditioned events.

三、研究方法

We start with a simple model first. Consider the waiting time sequence $W = \{W_n : n \geq 0\}$ of the GI/G/1 queue with traffic intensity $\rho < 1$. Let V_n $(n \geq 0)$ be the processing times, $U_n = A_n - A_{n-1}$ $(n \geq 1)$ the inter-arrival times, where A_n is the arrival time of the n-th customer. Set $X_n = V_{n-1} - U_n$, $n \geq 1$. Then W follows the well known Lindey recursion (see, e.g., Feller (1971)):

$$W_0 = 0, \quad \text{and} \quad W_n = (W_{n-1} + X_n)^+ \quad \text{for } n \geq 1, \tag{1}$$

where $a^+ = \max(a, 0)$.

Suppose we want to compute $\alpha(x) = P(W_\infty > x)$, where W_∞ is the steady-state waiting time distribution. If we apply importance sampling, the key quantity to be estimated is

$$\gamma_x = P(T(x) < \tau),$$

where $T(x) = \inf\{n \geq 0 : W_n > x\}$, $\tau = \inf\{n \geq 1 : W_n = 0\}$; see, in particular, § 5 of Glynn and Torres (1997).

Let $\psi(\theta) = \log E \exp(\theta X_i)$ be the cumulant moment generating function of X_i. We make the following assumptions:

A1 X is a non-lattice random variable.

A2 There exist $\theta^* > 0$ such that θ^* is the root of the equation $\psi(\theta) = 0$.

A3 $\psi(\theta) < \infty$ for $|\theta| < \theta_0$, $0 < \theta^* < \theta_0$.

Assume X has a common density function $f(\cdot)$. Then, under A1-A3, $f_\theta(x) = \exp(\theta x - \psi(\theta)) f(x)$ is also a density function. This specific change of measure is called "exponential twisting" or "exponential tilting." Let us denote $E(\cdot)$ and $P(\cdot)$ as the expectation and probability under $f(\cdot)$, and $E_\theta(\cdot)$ and $P_\theta(\cdot)$ as the expectation and probability under $f_\theta(\cdot)$.

Consider the indicator function

$$I(A) = 1 \quad \text{if } A \text{ occurs,} \quad \text{and} \quad I(A) = 0 \quad \text{otherwise.} \tag{2}$$

We estimate γ_x via "importance sampling" with twist θ by using an estimator based on replications of

$$Z_\theta(x) = \exp(-\theta S_{T(x)} + T(x)\psi(\theta)) I(T(x) < \tau) = L_{T(x)}(\theta) I(T(x) < \tau)$$

under P_θ, where $L_{T(x)}(\theta) = \exp(-\theta S_{T(x)} + T(x)\psi(\theta))$ is the likelihood ratio. By the convexity property of $\psi(\cdot)$ and the fact that $\psi'(0) = EX < 0$, we know $\psi'(\theta^*) > 0$. Hence $E_{\theta^*} X > 0$. Also, $P_\theta(T(x) < \infty) = 1$ for θ in a neighborhood of θ^*. Note that $L_{T(x)}(\theta^*)$ reduces to $\exp(-\theta^* S_{T(x)})$.

Results in Siegmund (1976) show that

$$\arg\min_\theta \mathrm{Var}(Z_\theta(x)) \to \theta^*, \quad \text{as } x \to \infty,$$

and, under suitable regularity conditions,

$$\mathrm{Var}(Z_{\theta^*}(x)) \sim c \cdot \exp(-2\theta^* x) - \gamma_x^2, \quad \text{as } x \to \infty, \tag{3}$$

for some constant c. (We write $a(x) \sim b(x)$ if $\lim_{x\to\infty} a(x)/b(x) = 1$.) Thus, θ^* is called the optimal twist and is a good value with which to do importance sampling. In fact, Lehtonen and Nyrhinen (1992) show that the exponential twist change of measure with parameter θ^* is the unique asymptotically optimal change of measure.

Now, we try to approximate θ^* via simulation. For this purpose, we need the following theorem due to Asmussen (1982).

Conditioned Limit Theorem for Random Walks

Let $\Rightarrow$ denote weak convergence. Then under A1-A3, there exists a random element $(\cdots, Y_{-1}, Y_0)$ of $\Re^{\{\cdots,-1,0\}}$ such that

$$P((\cdots, X_{T(x)-1}, X_{T(x)}) \in \cdot | T(x) < \tau) \Rightarrow P((\cdots, Y_{-1}, Y_0) \in \cdot),$$

as $x \to \infty$, where $(Y_{-n} : n \geq 0)$ has the property that

$$(\cdots, Y_{-n-1}, Y_{-n}) \Rightarrow (\cdots, Z_{-1}, Z_0)$$

as $n \to \infty$, with the Z_{-k} i.i.d. with common density function $f_{\theta^*}(\cdot)$.

To utilize the idea in the Theorem, let

$$\chi(x) = \frac{E_0[W_{T(x)}|T(x) < \tau]}{E_0[T(x)|T(x) < \tau]} = \frac{E_0[\sum_{i=1}^{T(x)} X_i|T(x) < \tau]}{E_0[T(x)|T(x) < \tau]}.$$

Above theorem suggests, when x is large, that

$$\chi(x) \approx E_{\theta^*}X = \psi'(\theta^*).$$

Now, a natural estimator for $\chi(x)$ is

$$\bar{\chi}_m(x) = \frac{\sum_{i=1}^m W_{T_i(x)}^{(i)} I(T_i(x) < \tau_i)}{\sum_{i=1}^m T_i(x) I(T_i(x) < \tau_i)}, \tag{4}$$

where $T_i(x), \tau_i$, and $W_{T_i(x)}^{(i)}$ are independent replications of $T(x), \tau$, and $W_{T(x)}$ respectively. The basic idea is that we approximate $\psi'(\theta^*)$ by $\bar{\chi}_m(\beta x)$, for some $\beta \in (0, 1]$ to be determined. If $\{T_i(x) < \tau_i\}$ is a rare event, then $\{T_i(\beta x) < \tau_i\}$ is also a rare event when $\beta > 1$. Therefore, we restrict β to be less than 1 in order to obtain an effective estimator of $\psi(\theta^*)$. (We say $\bar{\chi}_m(\cdot)$ is an "effective" estimate for $\chi(\cdot)$ if

$$\sum_{i=1}^m I(T_i(\cdot) < \tau_i) \geq c \tag{5}$$

for some constant c.)

we extend the idea above to rare event simulation of queueing networks. We describe the model detail of queueing networks we consider and the rare event simulation problem we would like to solve as follows.

Problem Description

We consider rare event simulations involving stable, single-class, open, generalized open Jackson networks. "Stable" means that the utilization of each queue is less than 1. "Single-class" means that there is only one type of customer. "Open" means that every arriving customer leaves the system with probability one.

A d-node queueing network with the following characteristics is a Jackson network (see Jackson (1957, 1963)):

1. Arrivals from the "outside" to node i follow a Poisson process with rate λ_i. That is, inter-arrival times from the "outside" to node i are independent and exponentially distributed with parameter λ_i.

2. Services times at node i are independent and exponentially distributed with parameter μ_i.

3. The probability that a customer who has completed service at node i will go to node j is P_{ij} (independent of the state of the system), where $i = 1, 2, \ldots, d, j = 0, 1, 2, \ldots, d$, and P_{i0} indicates the probability that a customer will depart from the system from node i.

A generalized Jackson network is a Jackson network with the following relaxation:

1. Inter-arrival times from the "outside" to node i are independent and identically distributed, but can be non-exponentially distributed.

2. Services times at node i are independent and identically distributed, but can be non-exponentially distributed.

We are interested in estimating the following performance measure associated with such networks:

$\pi_{\mathbf{K}} =$ the steady-state probability that the network population exceeds K.

Let us consider a d-node generalized Jackson network. Let $Q_i(\cdot)$, $A_i(\cdot)$, and $S_i(\cdot)$ be the processes associated with the queue-length, interarrival times, and service times of station i of the network. Specifically, we let $Q_i(t)$ be the number of customers in service and in queue i at time t, $A_i(\cdot)$ be the age process associated with the renewal exogenous arrival process (since inter-arrival times are independent and identically distributed, the arrival process is a renewal process; for general background on renewal and the associated age processes, see, e.g. § 5 of Karlin and Taylor (1975)) at node i, and $S_i(\cdot)$ be the age of the current customer that has been in service at node i if $Q_i(t) > 0$, and 0 otherwise. Then $X = \{X(t) : t \geq 0\}$ is a Markov process, where

$$X(t) = (Q_1(t), \ldots, Q_d(t), A_1(t), \ldots, A_d(t), S_1(t), \ldots, S_d(t)).$$

Let

$$Q(t) = \sum_{i=1}^{d} Q_i(t),$$

i.e., $Q(t)$ be the total network population of J at time t.

Let V be the set of states of X when $Q(t) = 0$ and define

$$t_0 = 0 \quad \text{and,} \quad t_i = \inf\{t > t_{i-1} : X(t) \in V, X(t^-) \notin V\} \quad \text{for } i \geq 1. \tag{6}$$

We define a V-cycle to be the process between two successive epochs t_i, t_{i+1} at which X enters the set V. We let ν be the steady-state distribution of X conditioned on X entering the set V, i.e.,

$$\nu(B) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(X(t_i) \in B) \quad \text{a.s.}$$

for any subset B of V. Let $N_V(t)$ be the number of V-cycles completed up to time t. We also let $Y = \int_{t_i}^{t_{i+1}} I(Q(s) \geq K)ds$, τ be the length of a V-cycle. Then assuming the limits below exists, it is evident that

$$\pi_K = \lim_{t\to\infty} \frac{\int_0^t I(Q(s) \geq K)ds}{t} = \lim_{t\to\infty} \frac{(\int_0^t I(Q(s) \geq K)ds)/N_V(t)}{t/N_V(t)} = \frac{E_\nu Y}{E_\nu \tau},$$

where the subscript ν in the expectation denotes the steady-state distribution of X at the beginning of an V-cycle.

We define

$$\begin{aligned} T_K &= \inf\{t \geq 0 : Q(t) \geq K\}, \\ T_0 &= \inf\{t \geq 0 : Q(t) = 0, Q(t^-) \neq 0\}. \end{aligned}$$

That is, T_K is the first time the network population exceeds K and T_0 is the first time the network population becomes zero.

Let us define

$$\gamma_K = P_\nu(T_K < T_0) \tag{7}$$

That is, γ_K is the probability that the system population reaches K during a busy period, given that the system starts empty. Then

$$\pi_K = \frac{\gamma_K \cdot E_\nu[\int_0^\tau I(X(s) \geq K)ds | T_K < T_0]}{E_\nu \tau}.$$

The technique used for estimating π_K is to simulate $Y = \int_0^\tau I(X(s) \geq K)ds$ under an appropriate change of measure, and simulate τ under the original dynamics. However, since X does not start from ν, one needs to apply some techniques that can deal with the initial transient problem in order to obtain accurate values of $E_\nu Y = E_\nu[\int_0^\tau I(X(s) \geq K)ds]$ and $E_\nu \tau$. The detailed description of this technique can be found in Nicola et al. (1993). But, it is true that the change of measure for estimating γ_K is closely related to the right change of measure for estimating Y (when K is big.) Specifically, suppose L is the likelihood ratio under which γ_K can be efficiently computed. Then

$$\begin{aligned} \text{Var}(YL) &= E(Y^2 L^2 | T_K < T_0)\gamma_K - \{E(YL|T_K < T_0)\gamma_K\}^2 \\ &= E\{E[Y^2|X(T_K)]L^2 \cdot I(T_K < T_0)\} - E\{E[Y|X(T_K)]L \cdot I(T_K < T_0)\}^2 \\ &= E\{\phi_2(X(T_K))L^2 \cdot I(T_K < T_0)\} - E\{\phi_1(X(T_K))L \cdot I(T_K < T_0)\}^2, \end{aligned}$$

where $\phi_1(\cdot) = E[Y^2|\cdot]$ and $\phi_2(\cdot) = E[Y|\cdot]$. Typically,

$$\phi_i(X(T_K)) = \theta(1), \quad i = 1, 2 \quad \text{as } K \to \infty.$$

(We say $a(n) = \theta(b(n))$ if there exist constants c_1, c_2 and n_0 such that

$$c_1 \cdot b(n) \leq a(n) \leq c_2 \cdot b(n)$$

for all $n \geq n_0$.) Then, we expect that

$$\text{Var}(YL) = \theta(\text{Var}(LI(T_K < T_0)).$$

In other words, $\text{Var}(YL)$ behaves in roughly the same way as the variance of the estimator of γ_K for K large. Hence, an efficient change of measure for estimating γ_K is the key to estimating π_K.

It is straightforward to estimate γ_K. Notice that a generalized Jackson network can be easily fit into a GSMP (Generalized Semi-Markov Process) framework. Therefore, so long as we can compute an appropriate importance sampling distribution, we can apply the importance sampling technique for GSMP described in Section 4 of Glynn and Iglehart (1989) to estimate γ_K.

The basic idea of obtaining an importance sampling distribution is as follows.

Simulate m sample paths from the original network starting with initial distribution ν and stop at $\min(T_{K'}, T_0)$, where $K' < K$.

Similar to (4), we let

$$\bar{U}_m^{(j)}(K') = \frac{\sum_{i=1}^m \sum_{l=1}^{N(i)} U_l^{(j)} I(T_i(K') < T_i(0))}{\sum_{i=1}^m N(i) I(T_i(K') < T_i(0))},$$

where $T_i(K')$ and $T_i(0)$ are independent replications of $T_{K'}$ and T_0, $U_l^{(j)}$ is independent replications of the interarrival time at node j, and $N(i)$ denotes the number of $U_l^{(j)}$ generated for sample path i. We then compute the importance sampling distribution for $U^{(j)}$ (interarrival time distribution at station j). We compute importance sampling distributions for $V^{(j)}$ (service time distribution at station j) and $M^{(j)}$ (routing distribution at station j) similarly.

四、結果與討論

We consider a sequence of rare events $\{A_K : K = 1, 2, \ldots\}$ ($\{A_x : x \geq 0\}$), and γ_K (γ_x) $\to 0$ denotes the probability of the rare event A_K (A_x). From the conditioned limit theorems of random walks (Asmussen, 1982; Iglehart, 1975), we usually can get an effective importance sampling distribution if we have samples conditioned on A_K (A_x), when K (x) large, does happen. The basic idea of our algorithm is to get importance sampling distributions via samples conditioned on a small K (x). For example, in the generalized Jackson network context, let A_K denote the event of the network population reach K before network becomes empty, the algorithm computes a good importance sampling distribution by simulating the original network for a small value of K, corresponding to "scaling down" the original rare event. We then simulate the desired rare-event probability, by "scaling up" the importance sampling distribution suggested by the key paths associated with the small K network. We call our proposed method SEEKPATH.

It is usually desirable to prove an importance sampling estimator is asymptotically optimal. For a definition and examples, see, e.g., p. 49 of Heidelberger (1995). In the queueing network context, the results of asymptotic optimal estimators is very limited, even for non-adaptive importance sampling estimators. To our best knowledge, the only rigorous proof of asymptotic optimality is for M/M/1 queues in tandem, and the result covers only for a certain parameter range of such networks (Glasserman and Kou, 1995).

Nevertheless, the empirical results indicate the SEEKPATH estimator enjoys the asymptotic optimal property for a much broader class of queueing models, and can be a pragmatic choice of practitioners.

The results of this project is a joint work with Professor Glynn at Stanford University. We will submit the results of this project to a suitable journal for publication in the near future.

參考文獻

Asmussen, Søren (1982), Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the GI/G/1 queue., *Adv. in Appl. Probab.*, 143–170.

Asmussen, Søren and Rubinstein, Reuven Y. (1995), Steady state rare events simulation in queueing models and its complexity properties, in *Advances in Queueing*, Probab. Stochastics Ser., 429–461, Boca Raton, FL: CRC.

Feller, William (1971), *An Introduction to Probability Theory and its Applications. Vol. II.*, New York: John Wiley & Sons Inc., second edition.

Frater, Michael R., Lennon, Tava M., and Anderson, Brian D. O. (1991), Optimally efficient estimation of the statistics of rare events in queueing networks, *IEEE Trans. Automat. Control*, 36(12), 1395–1405.

Frater, M.R. and Anderson, B.D.O. (1994), Fast simulation of buffer overflowss in tandem networks of GI/G/1 queues, *Annals of Oper. Res.*, 49, 207–220.

Glasserman, P. and Kou, S. (1995), Analysis of an importance sampling estimator for tandem queues, *ACM Trans. on Modeling and Computer Simulation*, 5, 22–42.

Glynn, P. W. and Iglehart, D. L. (1989), Importance sampling for stochastic simulations, *Manage. Sci.*, 35, 1367–1392.

Glynn, Peter W. and Torres, Marcelo (1997),Nonparametric estimation of tail probabilities for the single-server queue, in *Stochastic Networks: Stability and Rare Events*, 109–138, Springer-Verlag.

Hammersley, J. M. and Handscomb, D. C. (1965), *Monte Carlo Methods*, London: Methuen & Co. Ltd.

Heidelberger, P. (1995), Fast simulation of rare events in queueing and reliability models, *ACM Trans. on Modeling and Computer Simulation*, 5, 43–85.

Iglehart, Donald L. (1975), Conditioned limit theorems for random walks, in *Stochastic Processes and Related Topics*, 167–194, New York: Academic Press.

Jackson, J. R. (1957), Networks of waiting lines, *Oper. Res.*, 5, 518–521.

Jackson, J. R. (1963), Jobshop-like queueing systems, *Manage. Sci.*, 10, 131–142.

Karlin, Samuel and Taylor, Howard M. (1975), *A First Course in Stochastic Processes*, Academic Press, New York-London, second edition.

Lehtonen, T. and Nyrhinen, H. (1992), Simulating level-crossing probabilities by importance sampling, *Adv. in Appl. Probab.*, 24(4), 858–874.x

Nicola, V.F., Heidelberger, P., Shahabuddin, P., and Glynn, P.W. (1993), Fast simulation of steady-state availability in non-Markovian highly dependable systems, in *Proceedings of the Twnety Third International Symposium on Fault-Tolerance Computing*, 38–47.

Parekh, S. and Walrand, J. (1989), A quick simulation method for excessive backlogs in networks of queues, *IEEE Trans. Automatic Control*, 25, 54–66.

Shahabuddin, P. (1995), Rare event simulation in stochastic models, in *Proceedings of the 1995 Winter Simulation Conference*, 178–185, IEEE Computer Society Press.

Siegmund, D. (1976), Importance sampling in the Monte Carlo study of sequential tests, *Ann. Statist.*, 4(4), 673–684.