COMMUNICATIONS IN INFORMATION AND SYSTEMS
Vol. 3, No. 4, pp. 259-278, September 2004

002

PERFORMANCE ANALYSIS CONDITIONED ON RARE EVENTS: AN ADAPTIVE SIMULATION SCHEME*

V. S. BORKAR†, S. JUNEJA‡, AND A. A. KHERANI§

Dedicated to Sanjoy Mitter on the occasion of his 70th birthday.

Abstract. We consider the problem of simulation-based estimation of performance measures for a Markov chain conditioned on a rare event. The conditional law depends on the solution of a multiplicative Poisson equation. An adaptive scheme for learning the latter is proposed and analyzed. An example motivated by a well known problem in communication networks is given. Applications of the basic scheme to other related domains such as importance sampling for rare event simulation and the solution of a class of eigenvalue problems are also sketched.

Keywords: conditional performance analysis, multiplicative Poisson equation, stochastic approximation, adaptive simulation, reinforcement learning

1. Introduction.

Simulation-based performance analysis of discrete event systems such as queuing networks has traditionally concerned itself with estimation of probabilities or expectations associated with the underlying Markov chain. There are, however, situations where one is interested in *conditional* probabilities or expectations, conditioned on a 'rare' event of zero probability. (If the latter probability were positive, the problem could be reduced to that of estimating two expectations whose ratio gives the desired conditional expectation.) A typical scenario wherein such a situation might arise is when one wants to investigate how an extremal event in one part of a large interconnected system (such as overflow of a particular buffer or a burst of traffic in a queuing network) affects the functioning of another part of the system. Our aim here is to provide an adaptive scheme to handle such a situation when the conditioning event is a rare event of a particular kind, viz., that corresponding to a certain time average of a function of the underlying Markov chain exceeding a threshold larger than its mean. The conditional law of the chain conditioned on such an event turns out to be the law of another Markov chain with transition probabilities absolutely continuous w.r.t. the original transition probabilities. This change of measure depends on the solution to a *multiplicative Poisson equation*, which is in fact the problem of finding the Perron-Frobenius eigenfunction of a certain positive operator.

*Accepted for publication on January 28, 2004. Research supported in part by grant no. III.5(157)/99-ET from Dept. of Science and Technology, Govt. of India.

†School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Rd., Mumbai 400005, India, E-mail: borkar@tifr.res.in

‡School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Rd., Mumbai 400005, India, E-mail: juneja@tifr.res.in

§School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Rd., Mumbai 400005, India (now at INRIA, Sophia Antipolis, France), E-mail: Arzad-Alam.Kherani@sophia.inria.fr

This equation also arises in connection with dynamic programming for risk-sensitive control and its well-posedness as well as computational issues have been extensively studied in [10], [8], [9]. We use these developments to devise an adaptive scheme for on-line learning of this change of measure. This scheme uses for each simulation transition the current guess for the change of measure, which in turn uses the current guess for the solution to the multiplicative Poisson equation mentioned above. The latter is updated iteratively by a stochastic approximation scheme. This scheme is of the 'reinforcement learning' variety adapted from [8], [9]. (An adaptive importance sampling scheme based on reinforcement learning for estimating *expectations* associated with rare events was presented in [1].)

The conditional performance analysis problem described above is important in several contexts such as communication networks, supply chains, etc., and to our knowledge this is the first work to address it and provide a mathematical formulation as well as an algorithm to go with it. As it turns out, the key component of the algorithm is readily provided by existing developments in a completely different domain, viz., risk-sensitive control. To be precise, the core iteration of our two-tier algorithm is a 'linear' version of the reinforcement learning algorithm for risk-sensitive control. This is augmented by another iteration, a stochastic ascent scheme commonplace in stochastic approximation literature. The two are interlinked, but operate on different timescales which induce a 'leader-follower' behavior in the coupled scheme. This fits into the framework of two timescale stochastic approximation and can be analyzed as such. The overall scheme thus is a composition of existing ideas from other domains, a fact which helps us economize on the analysis presented here.

The paper is organized as follows: The next section describes the mathematical formalism underlying our scheme. Section 3 describes the algorithm and briefly sketches its convergence analysis. Section 4 lists two related problems which fit the general scheme, making them amenable to the techniques developed here. These are resp. the problem of asymptotically optimal adaptive importance sampling for rare event simulation and the problem of finding the Perron-Frobenius eigenvalue and eigenvector for a class of positive operators. Section 5 describes some simulation studies for a queuing example. Section 6 describes some important future directions for research.

2. Mathematical background. This section collects together the mathematical facts underlying our scheme. Some of it is standard folklore of the subject, but is repeated here nevertheless in order to make the account reasonably self-contained.

We shall consider an irreducible Markov chain $\{X_n, n \geq 0\}$ on a (*very large*) finite state space S. Denote by $\tilde{E}[\ \cdot\ ]$ its stationary expectations. Let $g : S \to \mathcal{R}$ and $\alpha > \tilde{E}[g(X_n)]$. For simplicity, we assume $\tilde{E}[g(X_n)] = 0$, which is equivalent to replacing α by $\alpha - \tilde{E}[g(X_n)]$ throughout in what follows. Formally, we are interested

in the behavior of this chain conditioned on the rare event

$$lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha,$$

or, to be precise, in the conditional transition probabilities

$$p^*(i,j) \stackrel{\Delta}{=} lim_{n\to\infty} P(X_1 = j | X_0 = i, \frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha),\ i,j \in S. \tag{1}$$

We shall do this by means of an adaptive scheme that may be viewed as 'conditional Monte Carlo'. For this purpose we need the *multiplicative Poisson equation*

$$V_\zeta(i) = \frac{e^{\zeta g(i)}}{\rho_\zeta}\sum_j p(i,j)V_\zeta(j),\ i \in S, \tag{2}$$

for a prescribed parameter $\zeta > 0$, where $p(\cdot,\cdot)$ are the original transition probabilities. Note that V_ζ, resp. ρ_ζ, are then the Perron-Frobenius eigenvector, resp. eigenvalue, of the positive operator

$$f(\cdot) \to e^{\zeta g(\cdot)}\sum_j p(\cdot,j)f(j). \tag{3}$$

LEMMA 1. For $\zeta > 0$, there exist $V_\zeta(\cdot) > 0, \rho_\zeta > 0$, satisfying (2), where $V_\zeta(\cdot)$ is unique up to multiplication by a positive scalar and ρ_ζ is the unique scalar given by

$$\ln \rho_\zeta = lim_{n\to\infty}\frac{1}{n}\ln E[e^{\sum_{m=0}^{n-1}\zeta g(X_m)}]. \tag{4}$$

See [2] for a proof. We fix the choice of V_ζ by imposing $V_\zeta(i_0) = \rho_\zeta$ for a prescribed $i_0 \in S$. The following is also standard, see, e.g., Lemma 6.4 of [22].

COROLLARY 1. The map $\zeta \to \rho_\zeta$ is convex and for any $\alpha > 0$ (more generally, $> \tilde{E}[g(X_n)]$, which is assumed to be zero here), there exists a unique $\zeta^* \stackrel{\Delta}{=} argmax_{\zeta\geq 0}(\zeta\alpha - \ell n(\rho_\zeta))$.

Let

$$\rho^* \stackrel{\Delta}{=} \rho_{\zeta^*},$$
$$V^* \stackrel{\Delta}{=} V_{\zeta^*},$$
$$\lambda^* \stackrel{\Delta}{=} \sqrt{\frac{\partial^2 \ell n(\rho_\zeta)}{\partial \zeta^2}\Big|_{\zeta=\zeta^*}}.$$

We also need the following Bahadur-Rao type exact asymptotic from Theorem 6.3 of [22] (see [25], [28] and [18, Chapter 9] for extensions and related results):

LEMMA 2. As $n \to \infty$,

$$P_x(\frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha_n) \sim \frac{V^*(x)e^{-n(\zeta^*\alpha - \ell n(\rho^*))}}{\zeta^*\sqrt{2\pi n\lambda^*}}e^{k\zeta^*},$$

where $\alpha_n = \alpha - k/n$ for a constant k, the subscript x on the l.h.s. stands for initial condition, and '$\sim$' stands for the fact that the ratio of the r.h.s. and the l.h.s. approaches one.

REMARK 1. We briefly sketch here how Lemma 2 may be derived using the proof techniques of [22]. To conform to the notation of [22], pp. 347-348, define

$$\begin{aligned}
\Lambda_n(\zeta) &\triangleq \frac{1}{n}\ell n\left(e^{\sum_{m=0}^{n-1} g(X_m)}\right), \\
\Lambda_n^*(c) &\triangleq sup_a(ac - \Lambda_n(a)), \\
\Lambda(\zeta) &\triangleq \rho_\zeta \text{ above}, \\
\Lambda^*(c) &\triangleq sup_a(ac - \Lambda(a)).
\end{aligned}$$

It may be noted that the suprema in the first and the third definition above are over a set $\{a : |a| < \bar{a}\}$ in [22], where $\bar{a} \in [0, \infty]$ is defined in terms of the constant arising in the Liapunov condition for geometric ergodicity (see p. 325 and 314-315 resp. of [22]) It is easy, however, to verify that in the finite state space set-up we are using, $\bar{a}$ may be taken to be ∞. For $n \geq 1$, let $a_n \in (0, \infty)$ denote the unique number such that $\Lambda_n'(a_n) = \alpha_n$ and $\Lambda_n^*(\alpha_n) = a_n\alpha_n - \Lambda_n(a_n)$ (i.e., $a_n = argmax_a(a\alpha_n - \Lambda_n(a))$). The difference with the corresponding definitions in [22], p. 348, is that they use $\alpha_n \equiv c \ \forall n$. Nevertheless, the fact that

$$\alpha_n - \alpha = -\frac{k}{n} = O(\frac{1}{n})$$

allows us to mimic the arguments of the key Lemma 6.4, p. 348-349, 359, of [22] to conclude that $a_n \to \zeta^*$ and

$$\Lambda_n^*(\alpha_n) = \Lambda^*(\alpha) - \frac{1}{n}\left(\ell n(V^*(x)) + k\zeta^*\right) + o(\frac{1}{n}).$$

(We observe here that ζ^*, V^* in our notation are precisely $a, \breve{f}$ in the notation of [22].) From this, our claim follows exactly along the lines of Theorem 6.3, p. 347-348, of [22].

This in turn leads to:

THEOREM 1. $p^*(\cdot,\cdot)$ is given by

$$p^*(i,j) = \frac{e^{\zeta^* g(i)} p(i,j) V^*(j)}{\rho^* V^*(i)}. \tag{5}$$

Proof. Note that

$$\begin{aligned} & P(X_1 = j | X_0 = i, \frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha) \\ = & \frac{P(X_0 = i, X_1 = j, \frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha)}{P(X_0 = i, \frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha)} \\ = & \frac{p(i,j)P_j(\frac{1}{n-1}\sum_{m=0}^{n-2} g(X_m) \geq \alpha - \frac{1}{n-1}(g(i) - \alpha))}{P_i(\frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha)} \\ \rightarrow & p^*(i,j) \end{aligned}$$

as defined above, in view of the preceding lemma. □

Note that this is nothing but a *Gibbs conditioning principle* (see, e.g., [13, section 3.3]) for Markov chains. The Markov property under the above conditional law can also be established by an analogous argument. In fact, one can show that for $m > 0$,

$$lim_{n\to\infty} P(X_m = x_m | X_k = x_k, 0 \leq k < m, \frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha) \rightarrow p^*(x_{m-1}, x_m).$$

Thus we conclude:

THEOREM 2. The regular conditional law of $\{X_m, m \geq 0\}$ conditioned on the event $\{X_0 = x, \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \geq \alpha\}$ converges to the law of a Markov chain starting at x with transition probabilities $p^*(\cdot,\cdot)$.

COROLLARY 2. For any $h : S \rightarrow \mathcal{R}$,

$$lim_{m\to\infty} lim_{n\to\infty} E[\frac{1}{m}\sum_{k=0}^{m-1} h(X_k) | \frac{1}{n}\sum_{k=0}^{n-1} g(X_m) \geq \alpha] = E^*[h(X_t^*)],$$

where $E^*[\,\cdot\,]$ is the stationary expectation for the irreducible Markov chain $\{X_n^*, n \geq 0\}$ with transition probabilities $p^*(\cdot,\cdot)$.

This result motivates our adaptive scheme, described in the next section. We conclude this section with a useful formula for $\partial \ell n(\rho_\zeta)/\partial\zeta$. Let $\eta^\zeta(\cdot)$ denote the unique stationary probability distribution for the irreducible Markov chain on S with transition probabilities

$$p^\zeta(i,j) \stackrel{\Delta}{=} \frac{e^{\zeta g(i)}}{V_\zeta(i)\rho_\zeta} p(i,j) V_\zeta(j), \;\; i,j \in S.$$

LEMMA 3. $\dfrac{\partial \ell n(\rho_\zeta)}{\partial \zeta} = \sum_i \eta^\zeta(i) g(i).$

For proof see, e.g., Proposition 4.9 of [22].

3. The adaptive algorithm. This section describes our algorithm and presents its convergence analysis. As the latter is mostly identical to that of [8], [9] with some ingredients from [6] and [7], we shall skip the technical details. These would be quite messy notationally *and* extremely lengthy. Instead we sketch the underlying arguments in broad strokes, with pointers to relevant literature as and when required.

Our adaptive scheme is based on a single simulation run $\{X_n\}$. Fix a distinguished state $i_0 \in S$. Let $\{a(n)\}, \{b(n)\}$ be positive scalar sequences satisfying

$$\sum_n a(n) = \sum_n b(n) = \infty, \ \sum_n (a(n)^2 + b(n)^2) < \infty, \ \frac{b(n)}{a(n)} \to 0. \tag{6}$$

These will serve as stepsizes or 'learning parameters' for our iterative scheme.

The algorithm

At each iteration n,

1. *simulate a transition from* $X_n = i$ *to* $X_{n+1} = j$ *(say) according to the current 'guess' of the transition probability* $p^*(i,j)$ *given by*

$$p_n(i,j) \stackrel{\Delta}{=} \frac{e^{\zeta_n g(i)}}{V_n(i)V_n(i_0)} p(i,j)V_n(j), \tag{7}$$

   normalized suitably to render it a probability vector, and

2. *update current guesses for* $(V^*(i), \zeta^*)$*, denoted by* $(V_n(i), \zeta_n)$*, according to the iterative scheme*

$$V_{n+1}(i) = V_n(i) + a(n) I\{X_n = i\} \Big[\frac{e^{\zeta_n g(i)}}{V_n(i_0)} V_n(X_{n+1}) \left(\frac{p(i, X_{n+1})}{p_n(i, X_{n+1})} \right) - V_n(i) \Big], \tag{8}$$

$$\zeta_{n+1} = \left(\zeta_n + b(n)\left(\alpha - g(X_{n+1})\right)\right)^+. \tag{9}$$

3. $n \to n+1$.

We now describe the rationale behind this scheme. The iterations are motivated by the following considerations: To solve the equation

$$V_\zeta(i) = \frac{e^{\zeta g(i)}}{\rho_\zeta} \sum_j p(i,j) V_\zeta(j), \ i \in S, \tag{10}$$

the following 'value iteration' scheme has been justified in [10]:

$$\tilde{V}^{n+1}(i) = e^{\zeta g(i)} \sum_j p(i,j) V^n(j),$$
$$V^{n+1}(i) = \tilde{V}^{n+1}(i) / \tilde{V}^{n+1}(i_0), \quad i \in S.$$

Recall that the solution of (10) is unique only up to a multiplicative scalar (see [10]), thus is rendered unique if the value of $V_\zeta(i)$ is fixed *a priori* for some choice of

i. The above scheme corresponds to fixing $V_\zeta(i_0) = 1$ and one correspondingly has $V^n(i_0) \to 1$. In [9], a 'reinforcement learning' version of this has been obtained in three steps, described below. While inspired by the above 'value iteration', it differs in one key aspect: It seeks the unique solution corresponding to $V_\zeta(i_0) = \rho_\zeta$.

1. Replace the above by a single iteration
$$V^{n+1}(i) = \frac{e^{\zeta g(i)}}{V^n(i_0)} \sum_j p(i,j) V^n(j), \ i \in S.$$
2. Replace the conditional average on the right by an actual evaluation at a simulated transition, i.e., by
$$\frac{e^{\zeta g(i)}}{V^n(i_0)} V^n(X_{n+1})$$
when $X_n = i$. This transition is executed with probability $p(i,\cdot)$.
3. Use the averaging property of stochastic approximation to get the desired average. This is achieved by making an incremental move in the direction suggested by the above, which leads to
$$V_{n+1}(i) = V_n(i) + a(n) I\{X_n = i\} \left(\frac{e^{\zeta g(i)}}{V_n(i_0)} V_n(X_{n+1}) - V_n(i) \right).$$
4. In our case, the simulated transition is part of a single simulation run for a Markov chain with time-varying transition probabilities $p_n(\cdot,\cdot)$ and *not* $p(\cdot,\cdot)$. Hence one 'unbiases' the r.h.s. above by multiplying the first term in the large parentheses with the likelihood ratio
$$\frac{p(X_n, X_{n+1})}{p_n(X_n, X_{n+1})},$$
which leads to (8) with $\zeta_n \equiv \zeta$.
5. In our case, ζ is replaced by a slowly varying sequence $\{\zeta_n\}$ given by (9). In view of Lemma 3, the iteration (9) is simply a projected stochastic gradient ascent algorithm for maximizing the function $\zeta \to \alpha\zeta - \ell n(\rho_\zeta)$, carried out on a slower timescale. The effect of the separation of timescales is that the iteration (8) sees $\{\zeta_n\}$ as 'quasi-static', permitting its analysis as though they were constant, as in [6].

Note that the $i-$th component gets updated only when $X_n = i$, which makes this an *asynchronous* stochastic approximation [7]. A *synchronous* version would be one in which all components are updated concurrently according to

$$V_{n+1}(i) = V_n(i) + a(n) \left(\frac{e^{\zeta_n g(i)}}{V_n(i_0)} V_n(X_{n+1}) \left(\frac{p(i, X_{n+1})}{p_n(i, X_{n+1})} \right) - V_n(i) \right). \tag{11}$$

Here X_{n+1} is a simulated random variable with law $p_n(i,\cdot)$, generated *separately* for each i.

If we 'freeze' $\zeta_n \equiv \zeta \ \forall n$, then this can be analyzed by the 'o.d.e.' approach to stochastic approximation as in [8], [9]. To be specific, one argues as in these articles that the iterates remain a.s. in the positive quadrant and bounded, whence they can be shown to asymptotically track the o.d.e.

$$\dot{x}_i(t) = \frac{e^{\zeta g(x_i(t))}}{x_{i_0}(t)} \sum_j p(i,j)x_j(t) - x_i(t), \ i \in S, \tag{12}$$

for $x(t) \stackrel{\Delta}{=} [x_1(t), \cdots, x_{|S|}(t)]^T$. It is then shown that this equation in the positive quadrant has a unique globally asymptotically stable equilibrium point, viz., the solution $\hat{V}_\zeta$ of the multiplicative Poisson equation (2) with $\hat{V}_\zeta(i_0) = \rho_\zeta$. In turn this implies by standard arguments of the 'o.d.e. approach' to stochastic approximation, that $V^n \to \hat{V}_\zeta$ a.s. In particular, $V^n(i_0) \to \rho_\zeta$ (*not* 1, as in the original value iteration of [10]). Recall that iteration (9) is the projected stochastic gradient scheme for maximizing $(\zeta\alpha - \ell n(\rho_\zeta))$ in view of Lemma 3. In order that (12) be justified, we need to run this stochastic gradient scheme on a slower timescale. This is achieved by the condition $b(n) = o(a(n))$ in (6) in view of the results of [6]. This allows us to 'freeze' $\zeta_n \equiv \zeta$ to analyze (8) as done above because $\{\zeta_n\}$ are changing on a slower timescale than (8) and thus are quasi-static. In turn, for the slower iterations (9), the fast iterations (8) are quasi-equilibrated. See [6] for the details of this argument.

This, however, justifies only the *synchronous* iterations. For the asynchronous iterations that we actually execute, the analysis of [7] implies that (12) should be replaced by

$$\dot{x}_i(t) = \Pi(t) \left(\frac{e^{\zeta g(x_i(t))}}{x_{i_0}(t)} \sum_j p(i,j)x_j(t) - x_i(t) \right), \ i \in S, \tag{13}$$

where $\Pi(t)$ is a diagonal matrix for each t with nonnegative diagonal elements adding up to 1 (i.e., they form a probability vector, reflecting in some sense the relative frequency with which the various components are being sampled - see [7]). Our 'irreducibility' hypotheses are sufficient to ensure that these diagonal elements remain bounded away from zero (see, e.g., the arguments on p. 843 of [7]). This implies that all components are being updated 'comparably often', as required for the analysis of [7]. If our simulated chain were asymptotically stationary, then $\Pi(t) \equiv \Pi_*$ for some time-invariant Π_*. (In fact, if the iterates converge, the resulting chain will indeed be asymptotically stationary.) (13) has the same equilibrium as (12). To see this, note that the r.h.s. of one vanishes at exactly the same point(s) as the other. What this implies is that *if* the iterations converge, they will converge to the desired limit. The proof that they *do* converge will require establishing that there are no other possible attractors other than equilibria. We do not have a proof of this at present for arbitrary $\Pi(\cdot)$ or Π^*, but our simulation results always showed convergence to an equilibrium.

An alternative would be to replace (8) by

$$V_{n+1}(i) = V_n(i) + a(\nu(i,n))I\{X_n = i\}\Big[\frac{e^{\zeta_n g(i)}}{V_n(i_0)}V_n(X_{n+1})\left(\frac{p(i,X_{n+1})}{p_n(i,X_{n+1})}\right) - V_n(i)\Big],$$

where

$$\nu(i,n) \triangleq \sum_{m=0}^{n-1} I\{X_m = i\},$$

is the 'empirical count' of state i at time n, for $i \in S$. Under some additional technical hypotheses which include additional conditions on $\{a(n)\}$, this can be shown to track the o.d.e. [7]

$$\dot{x}_i(t) = \frac{1}{|S|}\left(\frac{e^{\zeta g(x_i(t))}}{x_{i_0}(t)}\sum_j p(i,j)x_j(t) - x_i(t)\right),\ i \in S.$$

This is a time-scaled version of (12) and has identical phase portrait. Hence a.s. convergence follows as in the synchronous case. We tried this variation as well. It in fact gave faster convergence, though with higher fluctuation levels. This is as expected from the o.d.e. picture with the stepsize viewed as a discrete time step. In this version, the stepsize is decreased at a slower rate (because $\nu(i,n) < n$ in general) and thus the o.d.e. is 'simulated' on a faster timescale. At the same time, the fluctuations are larger because both the discretization error and error due to noise (which gets multiplied by the stepsize) increase with the stepsize. This trade-off is standard in stochastic approximation theory. Note also that the original motivation for using $a(\nu(i,n))$ instead of $a(n)$ in literature (see, e.g., [7]) is because $\{\nu(i,n)\}$ represents the local clock of the $i-$th processor in a distributed implementation in which each component is computed by a different processor. Here it is a centralized computation, so this becomes an artifice for sake of faster convergence with a theoretical guarantee, at the expense of the computational overheads to keep track of the $\nu(i,n)$'s and higher fluctuations.

4. Further applications. Here we briefly discuss two other important application domains where the above techniques can be usefully applied.

1. *Rare event simulation:* As is well known, estimating probabilities of rare events via naive simulation can be computationally prohibitively expensive. The importance sampling technique has been successful in efficiently simulating rare events for many stochastic systems (see, e.g., [16], [17]). It involves simulating the system under a change of measure and then unbiasing the resultant output by multiplying it with the likelihood ratio, i.e., the ratio of the original probability and the new probability of the generated sample

path. The challenge is to come up with a good change of measure so that the resultant output has significantly reduced variance compared to a naive estimator. It is well known that to estimate $P(A)$ for an event A, the zero variance estimator exists and under it the probability assigned to any event B equals $P(B|A)$. However, this zero variance estimator is typically not implementable as it requires a priori knowledge of $P(A)$, the unknown quantity. A good importance sampling distribution then is the one that closely approximates the zero variance estimator, i.e., the conditional measure, and is easy to identify and implement. Such distributions have been identified and successfully applied to simulate many queueing systems, reliability systems and other performance measures associated with random walks (see the references above).

For many such stochastic systems involving constituent processes that are Markov chains or more generally, Markov additive processes, the asymptotically optimal change of measure[1] involves finding the Perron-Frobenius eigenvalue and the corresponding eigenfunction of a member of a certain parameterized family of positive matrices (see, e.g., [12], [3], [23], [19] for such examples in single queues and queueing networks). This member is characterized by a parameter that is a unique solution to a specified equation, analogous to the problem considered in this paper. Thus, the methodology developed here proves useful in such rare event settings as well.

For example, in [11] it is shown that the asymptotically optimal change of measure to estimate the probability $P_x(\frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha)$ (using the notation from Section 2) is precisely the $p^*(\cdot,\cdot)$ specified by Theorem 1. An importance sampling estimator for this probability is the average of the i.i.d. samples of

$$I\{\frac{1}{n}\sum_{m=0}^{n-1} g(X_m) \geq \alpha\}\frac{p(X_0,X_1)p(X_1,X_2)\cdots p(X_{n-2},X_{n-1})}{p^*(X_0,X_1)p^*(X_1,X_2)\cdots p^*(X_{n-2},X_{n-1})}, \tag{14}$$

where $I\{A\}$ denotes the indicator of the event A and the ratio in the above equation is the unbiasing likelihood ratio.

As our experiments indicate, finding $p^*(\cdot,\cdot)$ purely by numerical techniques becomes difficult even for moderately large state spaces, making simulation a viable alternative. A natural way to implement our methodology is to first learn $p^*(\cdot,\cdot)$ via simulation to a specified degree of accuracy and then apply importance sampling as specified above. From a practical viewpoint an interesting question arises: What is the optimal level of accuracy to which

[1] Asymptotic optimality is a standard terminology in rare event simulation. Loosely speaking, under an asymptotically optimal change of measure the computational effort to estimate the rare event to a specified relative accuracy does not grow fast even as the event of interest becomes rarer.

$p^*(\cdot,\cdot)$ should be learnt so as to minimize the overall computational effort needed to estimate the probability of interest to a specified degree of accuracy? Too little effort in learning $p^*(\cdot,\cdot)$ may give a change of measure that has large variance in estimating the probability of interest. On the other hand, infinite amount of computational effort is needed to exactly learn $p^*(\cdot,\cdot)$ via simulation. We empirically observe this trade-off for a particular example in Section 5.

2. *Eigenvalue problems*

The proposed algorithm can be easily specialized to find the Perron-Frobenius eigenvalue and eigenvector for an irreducible non-negative matrix P, which may be taken to be substochastic w.l.o.g. This involves setting $\zeta_n =$ constant for all n and elimination of (9) from the algorithm.

The eigenvalue problems have diverse engineering and scientific applications. We refer the reader to [21], [14] and [15] for applications related to stability of dynamical systems, radiation transport systems, etc. These papers also propose adaptive techniques for determining eigenvalues of positive matrices similar to the one we encounter above. However, these rely on computationally expensive methodology that requires that many independent samples be generated from each state. Our approach based on updates using stochastic approximation offers a more economical alternative (see,e.g., [1]).

5. Application to a network of communication links. We now apply the foregoing to a concrete problem of great interest in communication networks. Consider a backbone network comprising a set of links $\mathcal{L}$ where each link $l \in \mathcal{L}$ has a capacity $C_l > 0$. A number of flows compete for access to these links, each flow is associated with a route identified by an ordered subset of $\mathcal{L}$. Write $l \in r$ when route r goes through link l. Let $\mathcal{R}$ denote the set of routes and let $N_r(t)$ denote the number of flows on route $r \in \mathcal{R}$ active at time instant t. Fairness considerations require that each of the $N_r(t)$ flows on route r have the same bandwidth $x_r(t)$ allocated to them. Call a bandwidth allocation $x(t) = \{x_r(t)\}_{r\in\mathcal{R}}$ feasible if it satisfies the capacity constraints,

$$\sum_{r:l\in r} N_r(t)x_r(t) \le C_l, \quad \forall l \in \mathcal{L}. \tag{15}$$

A general fairness criterion for allocating bandwidth to flows on different routes is introduced in [26]. It involves selecting a positive constant $\beta \neq 1$. For such a β, the fair allocation corresponds to the solution of the following optimization problem:

$$\text{Maximize} \sum_r N_r(t) \frac{x_r(t)^{1-\beta}}{1-\beta},$$

subject to the capacity constraints (15). Strict concavity of the function to be maximized ensures a unique allocation which is referred to as the $\beta-$proportionally fair

bandwidth allocation.

Two cases are of special interest: $\beta \to \infty$ implies the max-min fairness, i.e., the rate allocation which maximizes the minimum rate given to a flow in the network. On the other hand, $\beta \to 1$ gives the proportionally fair rate allocation, i.e., the rate allocation which maximizes the total throughput of the network (see [20]). For transfer of files controlled by Transport Control Protocol (TCP), it is argued in [20] that the rate achieved by the active file transfers on each route are proportionally fair. See [26] for further details.

In our experiment we consider the case where the flows on route r arrive according to a Poisson process of rate λ_r and terminate after receiving an exponentially distributed service (with mean μ_r, independent of the arrival process) that it brought with it. Let $N_r(t)$ denote the number of flows active at route r at time t. Allocation is determined dynamically as a function of $\{N_r(t)\}_{r\in\mathcal{R}}$ by solving the associated optimization problem. Note that the service rate obtained by a flow on route r, i.e., $x_r(t)$ is time varying since, for each r, $N_r(t)$ varies with time. Also note that the process $\{N_r(t)\}_{r\in\mathcal{R}}$ is a continuous time Markov chain.

This example is considered in [5], where it is shown that irrespective of the value of β, the Markov chain formed by the process $\{N_r(t)\}_{r\in\mathcal{R}}$ is stable under the simple condition

$$\sum_{r:l\in r} \lambda_r \mu_r \leq C_l, \quad \forall l \in \mathcal{L},$$

i.e., when the normalized offered load on each link is less than unity. It is also noted that finding the stationary distribution of the process $\{N_r(t)\}_{r\in\mathcal{R}}$ analytically is a non-trivial exercise and is known only for some specific networks for $\beta \to 1$ (see [5]). Thus development of techniques for efficient simulation of such networks is an important problem.

5.1. Simulating a Two Link Two Route Network. Figure 1 depicts a linear data network comprising two tandem links and two routes. This is a common scenario when a dial-up user is connected to a high speed backbone link via a single low speed access link. The backbone link carries traffic which is a superposition of the traffic generated by the dial-up user and the traffic from other source-destination pairs. The traffic from other source destination pairs is assumed to be clubbed together and forms Route 2 in the figure. A reason for choosing this particular network topology is the well known fact that for this network all $\beta-$proportional fair rates are equivalent.

Recall that $\{N_1(t), N_2(t)\}$ is a vector valued process corresponding to the number of ongoing transfers in the two routes at time t. Let $\{N_1^n, N_2^n\}$ be the discrete time Markov chain obtained by sampling the above process periodically at times placed δ units apart for δ small, i.e., $N_i^n = N_i(\delta n)$ for $i = 1, 2$ and all n. The transition probabilities for this Markov chain can be easily obtained from the proportionally fair

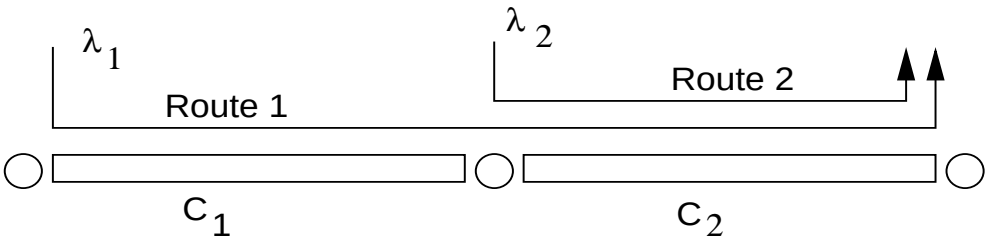

FIG. 1. *Two Link Two Route Network*

rates. It is easily shown that for this network,

$$x_1(t) = \begin{cases} \frac{C_1}{N_1(t)} & \text{if } \frac{C_2}{N_1(t)+N_2(t)} > \frac{C_1}{N_1(t)}, \\ \\ \frac{C_2 N_1(t)}{N_1(t)+N_2(t)} & \text{otherwise,} \end{cases}$$

$$x_2(t) = \frac{C_2 - N_1(t)x_1(t)}{N_2(t)}.$$

5.1.1. Estimating the conditional expectation. We first estimate via simulation, the expected number of flows in steady state at Route 1 conditioned on the event that Route 2 is heavily loaded for a long period of time. Specifically, we estimate:

$$\lim_{m\to\infty} \lim_{n\to\infty} E[\frac{1}{m}\sum_{i=1}^{m} N_1^i \,|\, \frac{1}{n}\sum_{i=1}^{n} I\{N_2^i > N\} \geq \alpha] = E^*[N_1^i],$$

where $E^*[\cdot]$ is the stationary expectation for the Markov chain $\{N_1^n, N_2^n\}$ under $p^*(\cdot,\cdot)$ as specified in Theorem 2 and N and α are constants specified later. In our simulations we estimate $E^*[N_1^i]$ simultaneously while learning $p^*(\cdot,\cdot)$ using the proposed adaptive algorithm with $g(N_1^n, N_2^n) = I\{N_2^n > N\}$.

The parameters used in the simulation are: $\lambda_1 = 0.2$, $\lambda_2 = 0.5$, $C_1 = 0.5$, $C_2 = 1.0$, $\mu_1 = \mu_2 = 1.0$, $N = 5$, $\delta = 0.5$ and the simulation is started with an initial value of $\zeta = 0.0$. These choices are made so as to facilitate the numerical validation; see section 5.1.2. We set $\alpha = 0.2463$ (from simulations we observed that under the original measure $P(N_2^i > N)$ in steady state equals 0.0463, so α is greater than this). We restrict the state space of each N_i^n to be less than or equal to 300 so that the resultant state space is finite. (In our simulations these bounds are never achieved.) We use step-sizes $a(n) = \frac{0.15}{\lceil n*10^{-6} \rceil}$ and $b(n) = \frac{a(n)}{\ln(n)*10^3}$. In our experiments not reported here, we found similar convergence rates with other related pairs of step-sizes.

Figure 2 shows the value of $\{\zeta_n\}$ as obtained during the simulation as a function of the number of iterations of the algorithm. It is seen from Figure 2 that this converges to the value obtained numerically in Section 5.1.2.

5.1.2. Numerical validation. We validate the results obtained via simulation by those computed using numerical techniques using *matlab* (*matlab* uses the QR algorithm to find the eigenvalues of a matrix; see, e.g, [24]). Numerical techniques have far greater storage requirement as storing the transition matrix requires $O(|S|^2)$ memory, where $|S|$ denotes the state space size. Note that our simulation methodology requires $O(|S|)$ storage. In our experiment $|S| = 90,000$. Storing a matrix of size 8.1×10^9 is not viable, thus to implement the numerical techniques we truncated the state space to $[0, 25] \times [0, 25]$ (now $|S| = 625$) and the arrival and service rates are chosen so that the probability of the number of flows in any link exceeding 25 is extremely small.

Implementing any numerical procedure requires selecting an initial value of ζ_0, then numerically computing the Perron-Frobenius eigenvalue ρ_ζ and/or related quantities such as its derivative with respect to ζ to update the value of ζ_n, and repeating this procedure till convergence to ζ^* is observed. The resulting plot of $\alpha\zeta - \ln \rho_\zeta$ vs. ζ for $\alpha = 0.2463$ is shown in Figure 3. It is seen from Figure 3 that the maximum of the function $\alpha\zeta - \ln \rho_\zeta$ is achieved for ζ near and to the right of $\zeta = 0.008$.

In Figure 4 we plot $\frac{1}{n}\sum_{i=1}^{n} N_1^i$ vs. n obtained from simulation. This is an estimate of the conditional mean number of flows active in route 1. The conditional value of $E^*[N_1^i]$ obtained from simulations is around 1.24 while it can be shown that the unconditional mean of N_1^i in steady state is 0.85.

Figure 5 shows the values of $\{\zeta_n\}$ obtained from the simulation with the same system parameters as used in previous reported simulations. The only difference now is that the updates of the transition probabilities for any state (N_1, N_2) are now based on the step sizes $a(\nu((N_1(n), N_2(n)), n))$, while the update of ζ uses a fixed state independent sequence $\{b(n)\}$. ζ is seen to converge to the correct value in this case as well.

5.1.3. Estimating the rare event probability. In order to demonstrate the utility of the proposed methodology to rare event simulation we also estimate

$$\hat{p}_n := P\left(\frac{1}{n}\sum_{i=1}^{n} I(N_2(i) > N) \geq \alpha\right)$$

for $N = 5$, $\alpha = 0.9$ and $n = 100$.

As we discussed in Section 4, there may exist an optimal level of accuracy to which $p^*(\cdot, \cdot)$ should be estimated so that the overall computational effort to estimate $\hat{p}_n$ to the specified degree of accuracy is minimized. We observe this trade-off empirically. The accuracy of estimation of $p^*(\cdot, \cdot)$ is measured by observing the stability of the estimate of ρ_ζ. For a fixed $\epsilon > 0$, let k_ϵ be the transition number at which estimate of ρ_ζ converged to within ϵ of its previous value 100 times in succession. Since we have $V_\zeta(0, 0) = \rho_\zeta$, this is effectively also a convergence criterion in terms of the estimates of $V_\zeta(0, 0)$. We run the algorithm for different values of ϵ. For each such ϵ, we obtain

ϵ	k_ϵ	Estimated $\hat{p}_N$ $(\times 10^5)$	T_ϵ
$1.0e-04$	$1.1e+04$	1.337773	$4.0e+10$
$1.0e-05$	$5.7e+06$	1.318673	$1.7e+09$
$1.0e-06$	$1.5e+08$	1.314035	$4.6e+09$
$1.0e-07$	$5.1e+09$	1.316046	$1.9e+10$
Naive	0	1.340516	$4.6e+10$

TABLE 1

This table shows the the estimated value of the rare event probability $\hat{p}_n$, when the optimal importance sampling change of measure is learnt to varying degrees of accuracy.

an estimate $p^*_\epsilon(\cdot,\cdot)$ of $p^*(\cdot,\cdot)$. This is then used to generate i.i.d. samples as shown in (14). From these samples we iteratively determine the sample mean and sample variance. The simulation stops when the 95% confidence interval of the probability, derived using standard methodology based on the central limit theorem along with the estimated mean and variance (see, e.g., [17], [19]), lies within 5% of its estimated mean value. Let T_ϵ denote the total computational effort, as measured by the number of transitions of the Markov chain generated via simulation till the desired accuracy level is attained.

In Table 1 we list the values of estimate of $\hat{p}_n$ for different ϵ. Also shown are the values of k_ϵ, and T_ϵ. We use constant step sizes $a = 0.15$ and $b = 10^{-6}$ for finding the change of measure. The values of network parameters are the same as those used in simulations described in Section 5.1.1. The estimates of $\hat{p}_n$ obtained via naive simulation are also shown.

From Table 1 we observe that the minimum computational effort corresponds to $\epsilon = 1.0e{-}05$, where we get about thirty times improvement over naive simulation. However, this comparison based on number of transitions has a minor flaw that it doesn't account for the differential of computational effort per transition in the two methodologies (i.e., proposed and naive). In the proposed methodology, in the initial stages of the simulation where the change of measure is adaptively learnt, additional effort at each transition is needed to update the estimates of (V^*, ζ^*). Once the change of measure is learnt and fixed, the effort required using importance sampling differs from naive simulation only in that the likelihood ratio needs to be updated at each transition. As the data in Table 1 indicates, when ϵ equals $1.0e{-}05$, the adaptive learning phase lasts for only 1/300 of the total number of transitions. Thus, the comparison based on number of transitions is reasonably accurate.

6. Future directions. In conclusion we point out some important topics for future work in this direction.

1. A natural extension of this framework would be to develop schemes for sim-

ulating conditional expectations conditioned on more than one event of the above type, i.e., on

$$\text{`}\lim_{n\to\infty}\text{'}\,\frac{1}{n}\sum_{m=0}^{n-1} g_i(X_m) \geq \alpha_i$$

for $1 \leq i \leq N, N > 1$. Let $\alpha \stackrel{\Delta}{=} [\alpha_1, \cdots, \alpha_N]$, $g(\cdot) \stackrel{\Delta}{=} [g_1(\cdot), \cdots, g_N(\cdot)]$ and $\zeta \stackrel{\Delta}{=} [\zeta_1, \cdots, \zeta_N] \in (\mathcal{R}^N)^+$. Then the results of [28] indicate that all one needs is to repeat the foregoing with $\zeta g(\cdot)$, $\zeta\alpha$ replaced by resp. $\langle \zeta, g(\cdot)\rangle$, $\langle \zeta, \alpha\rangle$. Unfortunately these results have been derived under a strong assumption equivalent to assuming that $p(i,j) > 0\ \forall i,j$, which does not hold, e.g., for queuing systems. Thus they need to be extended appropriately to more general chains.

2. For very large state spaces, as indeed will be the case in many important applications such as large networks of queues, the curse of dimensionality will make the 'ideal' scheme proposed above impractical. This calls for an additional layer of approximation. Taking cue from reinforcement learning literature [4], one may then approximate V^* by an element of a parametrized family of functions, e.g., linear combinations of a fixed finite basis of functions with the weights as 'parameters' to be tuned. The understanding here is that the dimensionality of the parameter space is much smaller than $|S|$. One then hopes to write a recursive scheme for learning the optimal parameters within the given search space. See [4] for such approaches in the more 'classical' reinforcement learning framework. These need to be extended to the schemes for 'risk-sensitive control' invoked here.

REFERENCES

[1] T. P. I. Ahamed, V. S. Borkar, and S. Juneja, *Adaptive importance sampling technique for Markov chains using stochastic approximation*, submitted.

[2] S. Balaji and S. P. Meyn, *Multiplicative ergodicity and large deviations for an irreducible Markov chain*, Stochastic Processes and Their Appl., 90(2000), pp. 123-144.

[3] B. Beck, A. R. Dabrowski, and D. R. Mcdonald, *A unified approach to fast teller queues and ATM*, Advances in Applied Probability, 31(1999), pp. 758-787.

[4] D. Bertsekas and J. N. Tsitsiklis, *Neurodynamic Programming*, Athena Scientific, Belmont, Mass., 1997.

[5] T. Bonald, L. Massoulie, *Impact of fairness on Internet performance*, SIGMETRICS'01, 2001.

[6] V. S. Borkar, *Stochastic approximation with two time scales*, Systems and Control Letters, 29(1997), pp. 291-294.

[7] V. S. Borkar, *Asynchronous stochastic approximations*, SIAM J. Control and Optimization, 36(1998), pp. 840-851. Erratum in ibid., 38(2000), pp. 662-663.

[8] V. S. Borkar, *Q-learning for risk-sensitive control*, Mathematics of Operations Research, 27(2002), pp. 294-311.

[9] V. S. Borkar, *A sensitivity formula for risk-sensitive cost and the actor-critic algorithm*, Systems and Control Letters, 44(2001), pp. 339-346.

[10] V. S. Borkar and S. P. Meyn, *Risk-sensitive optimal control for Markov decision processes with monotone cost*, Mathematics of Operations Research, 27(2002), pp. 192-209.

[11] J. Bucklew, *Large Deviations Techniques in Decision, Simulation and Estimation*, John Wiley, New York, 1990.

[12] C. S. Chang, P. Heidelberger, S. Juneja, and P. Shahabuddin, *Effective bandwidth and fast simulation of ATM intree networks*, Performance Evaluation, 20(1994), pp. 45-65.

[13] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*(2nd ed.), Springer Verlag, New York, 1998.

[14] P. DESAI, *Adaptive Monte Carlo Methods for Solving Eigenvalue Problems*, Ph.D. Dissertation, Dept. of Management Science and Engineering, Stanford Uni., CA, March 2001.

[15] P. Y. Desai and P. W. Glynn, *A Markov chain perspective on adaptive Monte Carlo algorithms*, Proceedings of 2000 Winter Simulation Conference, 2001, pp. 379-384.

[16] P. W. Glynn and D. L. Iglehart, *Importance sampling for stochastic simulations*, Management Science, 35(1989), pp. 1367-1392.

[17] P. Heidelberger, *Fast simulation of rare events in queuing and reliability models*, ACM Transactions on Modeling and Computer Simulation, 5(1995), pp. 43-85.

[18] J. L. Jensen, *Saddlepoint Approximations*, Oxford Uni. Press, Oxford, UK, 1995.

[19] S. Juneja, 2001. *Importance sampling and the cyclic approach*, Operations Research, 49:6(2001), pp. 900-912.

[20] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, *Rate control for communication networks: shadow prices, proportional fairness and stability*, Operations Research, 49(1998), pp. 237-252.

[21] C. Kollman, K. Baggerly, D. Cox, and R. Picard, *Adaptive importance sampling on discrete Markov chains*, Annals of Applied Probability, 9(1999), pp. 391-412.

[22] I. Kontoyiannis and S. P. Meyn, *Spectral theory and limit theorems for geometrically ergodic Markov processes*, Annals of Applied Probability, 13(2003), pp. 304-362.

[23] D. P. Kroese and V. F. Nicola, *Efficient simulation of a tandem Jackson network*, Proc. of the 1999 Winter Simulation Conference, IEEE Press, pp. 411-419.

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd edition, Johns Hopkins, Baltimore. 1996.

[25] H. D. Miller, *A convexity property in the theory of random variables defined on a finite Markov chain*, The Annals of Mathematical Statistics, 32(1961), pp. 1260-1270.

[26] J. Mo and J. Walrand, *Fair end-to-end window-based congestion control*, IEEE/ACM Transactions on Networking, 8(2000), pp. 556-567.

[27] S. M. Ross, *Simulation* (2nd ed.), Academic Press, San Diego, 1997.

[28] C. Schroeder, *I-projection and limit theorems for discrete parameter Markov processes*, The Annals of Probability, 21(1993), pp. 759-790.

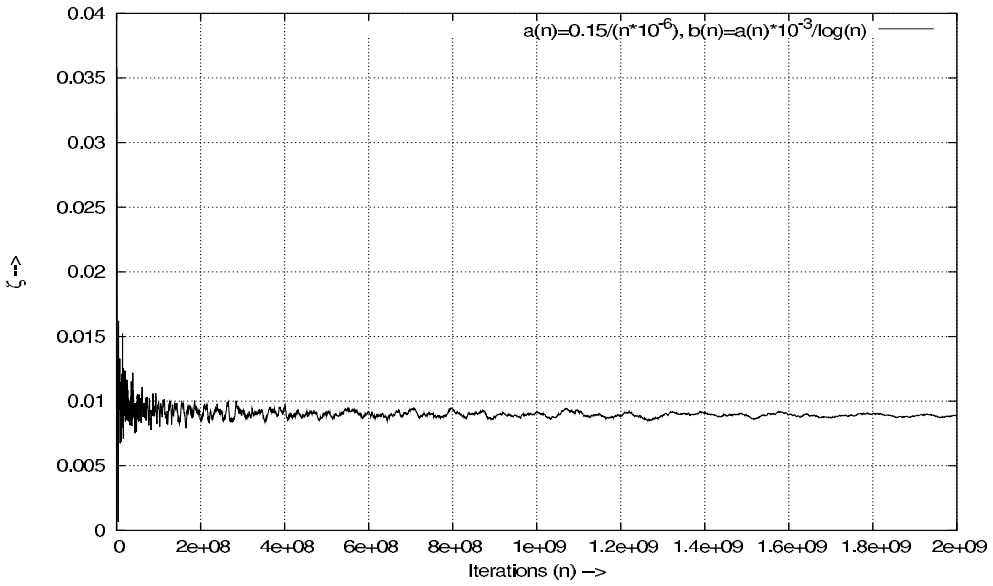

FIG. 2. *Values of ζ obtained from the simulation (state independent step sizes).*

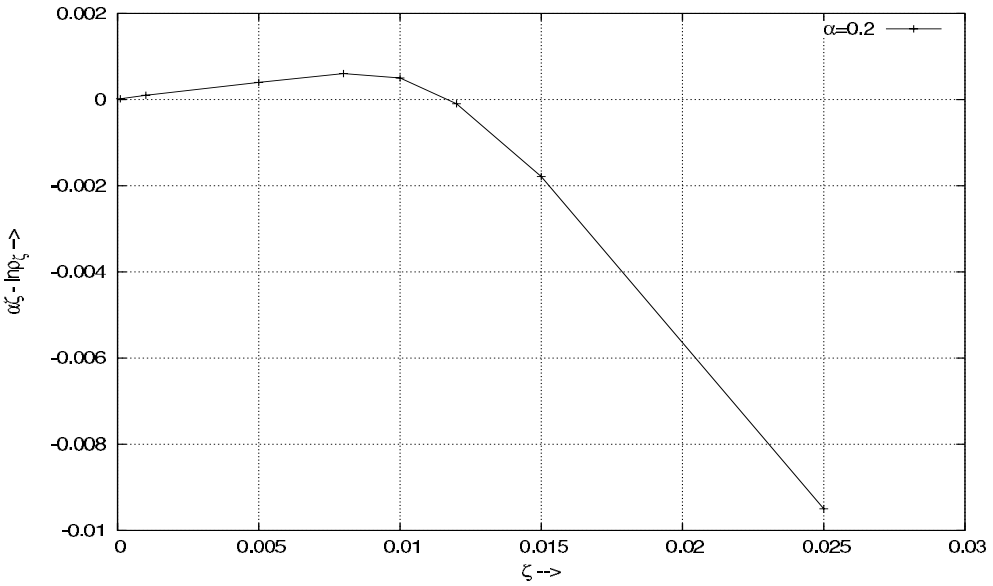

FIG. 3. *$\alpha\zeta - \ln \rho_\zeta$ computed using numerical methods.*

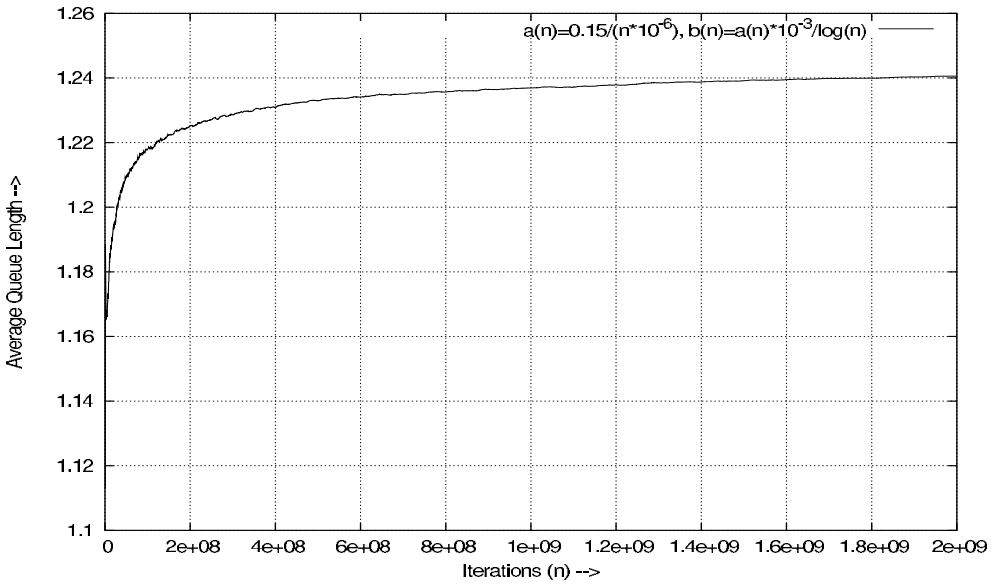

FIG. 4. *Average queue length* $\frac{1}{n}\sum_{i=1}^{n} N_1(i)$ *vs.* n.

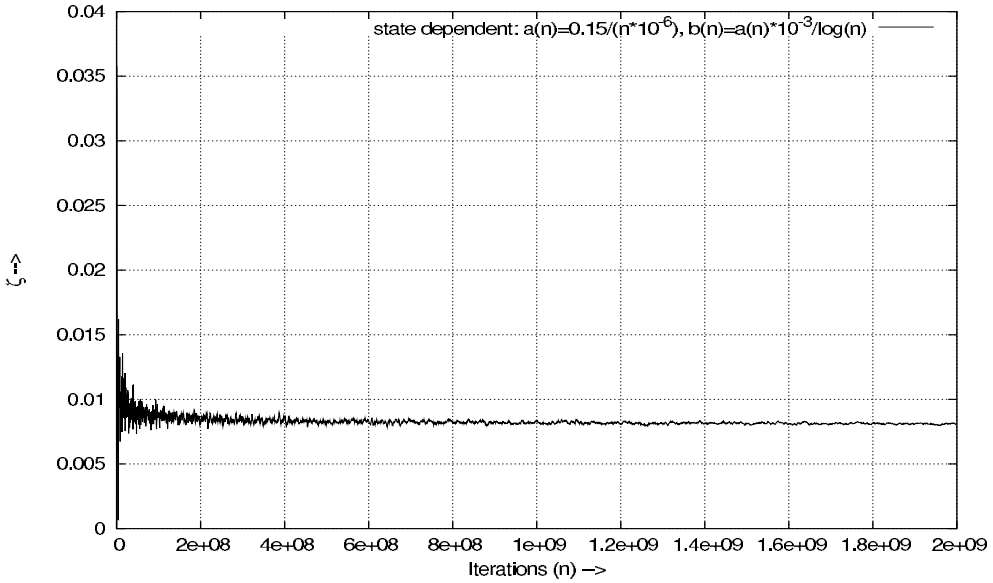

FIG. 5. *Values of* ζ *obtained from the simulation (state dependent step sizes).*