EXPERIMENTAL RESULTS FOR GRADIENT ESTIMATION AND OPTIMIZATION OF A MARKOV CHAIN IN STEADY-STATE

Pierre L'Ecuyer
Département d'I.R.O., Université de Montréal,
C.P. 6128, Montréal, H3C 3J7, Canada

Nataly Giroux
Département d'informatique, Université Laval,
Ste-Foy, Québec, G1K 7P4, Canada

Peter W. Glynn
Operations Research Department, Stanford University,
Stanford, CA94305, U.S.A.

ABSTRACT. Infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method have drawn lots of attention recently, as ways of estimating the gradient of a performance measure with respect to continuous parameters in dynamic stochastic systems. In this paper, we experiment with the use of these estimators in stochastic approximation algorithms, to perform so-called "single-run optimizations" of steady-state systems, as suggested in [23]. We also compare them to finite-difference estimators, with and without common random numbers. In most cases, the simulation length must be increased from iteration to iteration, otherwise the algorithm converges to the wrong value. We have performed extensive numerical experiments with a simple M/M/1 queue. We state convergence results, but do not give the proofs. The proofs are given in [14].

1. THE MODEL AND THE STOCHASTIC APPROXIMATION APPROACH

We consider a Markov chain $\{X_i(\theta,s,\omega), i = 0,1,\ldots\}$ with Borel state space S, defined over a probability space $(\Omega,\Sigma,P_{\theta,s})$. Its associated probability measure $P_{\theta,s}$ depends on the parameter vector θ, where $\theta \in \bar{\Theta} \subset \mathbb{R}^d$, and on the (deterministic) initial state $X_0(\theta,s,\omega) = s \in S$. A cost $f(\theta,x)$ is incurred whenever we visit state x (except for the initial state X_0), for measurable $f : \bar{\Theta} \times S \to \mathbb{R}$. Let

$$h_t(\theta,s,\omega) = \frac{1}{t}\sum_{i=1}^{t} f(\theta, X_i(\theta,s,\omega)) \tag{1}$$

be the average cost for the first t steps and

$$\alpha_t(\theta,s) = \int_\Omega h_t(\theta,s,\omega)\mathrm{d}P_{\theta,s}(\omega), \tag{2}$$

its expectation. Let $\bar{S}$ be a subset of S which can be viewed as the set of admissible initial states. For example, $\bar{S}$ can consist of a single state s_0 if all simulation subruns (below) are started from that state. In most other cases, $\bar{S}$ will be a compact subset of S. We assume that

$$\lim_{t\to\infty} \sup_{\theta\in\Theta, s\in\bar{S}} |\alpha_t(\theta,s) - \alpha(\theta)| = 0, \tag{3}$$

where $\alpha(\theta)$ is the steady-state average cost for running the system at parameter level θ, and that

$$\lim_{t\to\infty} \sup_{\theta\in\Theta, s\in\bar{S}} |\nabla_\theta \alpha_t(\theta,s) - \nabla_\theta \alpha(\theta)| = 0. \tag{4}$$

We assume that each $\alpha_t(\cdot,s)$ and α are differentiable. Suppose one is interested in minimizing $\alpha(\theta)$ over Θ, a compact and convex subset of $\bar{\Theta}$ such that each point of Θ has a neighborhood inside $\bar{\Theta}$. Let $\theta^* \in \Theta$ be the optimum (assumed to be unique). We consider a stochastic approximation (SA) algorithm of the form

$$\theta_{n+1} := \pi_\Theta(\theta_n - \gamma_n Y_n). \tag{5}$$

for $n \geq 1$, where θ_n is the parameter value at the beginning of iteration n ($\theta_1 \in \Theta$ is random with known distribution), Y_n is an estimate of the gradient $\nabla\alpha(\theta_n)$ obtained at iteration n, $\{\gamma_n, n \geq 1\}$ is a (deterministic) positive sequence decreasing to 0 and such that $\sum_{n=1}^{\infty}\gamma_n = \infty$, and π_Θ denotes the projection on the set Θ. In what follows, except when stated otherwise, we will assume that $\gamma_n = \gamma_0 n^{-1}$ for some constant $\gamma_0 > 0$.

Each Y_n is obtained by simulating the system for one or more subrun(s) of finite duration. Each simulation subrun corresponds essentially to one copy of the Markov chain described above, with initial state $s \in \bar{S}$. Specific ways of obtaining Y_n can be based, for example, on IPA, LR, or finite differences. We recall them in the next section. In general, since Y_n must be obtained in finite time, it is a biased estimator of $\nabla\alpha(\theta)$.

Let $s_n \in \bar{S}$ denote the *state* of the system at the beginning of iteration n. In the simulation program, this corresponds to the description of all the objects in the system, event list, etc.. We assume that when θ_n and s_n are fixed, the distribution of (Y_n, s_{n+1}) is completely specified and independent of the past iterations (but may depend on n). Here, Y_0 is a dummy value. Denote by $E_n(\cdot)$ the conditional expectation $E(\cdot \mid \theta_n, s_n)$, that is the expectation conditional on what is known at the beginning of iteration n. Assume that Y_n is integrable for all $n \geq 1$ and let:

$$Y_n = \nabla\alpha(\theta_n) + \beta_n + \epsilon_n \tag{6}$$

where $\beta_n = E_n[Y_n] - \nabla\alpha(\theta_n)$ represents the (conditional) bias on Y_n given (θ_n, s_n), while ϵ_n is a random noise, with $E_n(\epsilon_n) = 0$. The next proposition gives simplified sufficient conditions for the convergence of (5) to an optimum. More general conditions are given, e.g., in [10, 11, 14, 16, 18].

PROPOSITION 1. *Let $d = 1$. Θ is now a closed interval of $\mathbb{R}$. Assume that $\alpha(\theta)$ is strictly unimodal over $\bar{\Theta}$. If $\lim_{n\to\infty}\beta_n = 0$ and $\sum_{n=1}^{\infty} E_1(\epsilon_n^2)n^{-2} < \infty$ with probability one, then $\lim_{n\to\infty}\theta_n = \theta^*$ with probability one.* ∎

2. GRADIENT ESTIMATORS

2.1. Finite differences (FD)

Here, we consider *central* FD. For other variants, like *forward* FD, see [6, 10, 18]. For simplicity, let $d = 1$. Take a deterministic positive sequence $\{c_n, n \geq 1\}$ that converges to 0. At iteration n, simulate from some initial state $s_n^- \in \bar{S}$ at parameter value $\theta_n^- = \pi_\Theta(\theta_n - c_n)$ for t_n transitions. Simulate also (independently) from state $s_n^+ \in \bar{S}$ at parameter value $\theta_n^+ = \pi_\Theta(\theta_n + c_n)$ for t_n transitions. Let ω_n^- and ω_n^+ denote the respective sample points. The FD gradient estimator is

$$Y_n = \frac{h_{t_n}(\theta_n^+, s_n^+, \omega_n^+) - h_{t_n}(\theta_n^-, s_n^-, \omega_n^-)}{\theta_n^+ - \theta_n^-}. \tag{7}$$

For $d > 1$, just repeat the procedure for each component of θ_n.

Here, there are different sources of bias: there is bias due to the fact that we just simulate over a finite horizon, bias due the finite differences, and bias due to the possibly different initial states s_n^- and s_n^+. To get $\beta_n \to 0$, one typically needs to take $t_n \to \infty$. In the light of [6, 10], where related problems are discussed, reasonable choices for the sequences might be for instance $t_n = t_a + t_b n$ and $c_n = c_0 n^{-1/6}$ for appropriate constants t_a, t_b, and c_0. One simple way to choose the initial states of the subruns is as follows. Start the first subrun of iteration n from state $s_n \in \tilde{S}$, then take the terminal state of any given subrun as the initial state of the next one. (Project on $\tilde{S}$ whenever necessary.) For s_{n+1}, take the terminal state of the last subrun of iteration n. Another way is to take the same initial state for each subrun: $s_n^- = s_n^+ = s_n$. One can also take (reset) $s_n = s_0$ for all n, for a fixed state s_0. In any case, this method is usually plagued by a huge variance on Y_n, which makes it converge very slowly, at least when the subruns are performed with "independent" random numbers.

2.2. Finite differences with common random numbers (FDC)

One way to reduce the variance in FD is to use common random numbers across the subruns at each iteration, start all the subruns from the same state: $s_n^- = s_n^+ = s_n$, and synchronize. More specifically, one views ω as representing a sequence of $U(0,1)$ variates, so that all the dependency on (θ, s) appears in $h_t(\theta, s, \cdot)$. Take $\omega_n^+ = \omega_n^- = \omega_n$. Since the subruns are aimed at comparing very similar systems, $h_{t_n}(\theta_n^+, s_n, \omega_n)$ and $h_{t_n}(\theta_n^-, s_n, \omega_n)$ should be highly correlated, especially when c_n is small, so that considerable variance reductions might be obtained. Conditions that *guarantee* variance reductions are given in [3, 18]. In practice, this method is not always easy to implement. See the discussion in [14]. Reasonable choices for the sequences are $t_n = t_a + t_b n$ and $c_n = c_0 n^{-1/5}$ for appropriate constants t_a, t_b, and c_0. This is somewhat justified by the results of [6], where the related finite-horizon gradient estimation problem is analyzed.

2.3. A likelihood ratio (LR) approach

With the LR approach [1, 5, 7, 12, 17, 19, 20], to differentiate the expectation (2) with respect to θ, we first take a probability measure G independent of θ that dominates the $P_{\theta,s}$'s for $\theta \in \bar{\Theta}$, $s \in \tilde{S}$, and rewrite:

$$\alpha_t(\theta, s) = \int_\Omega h_t(\theta, s, \omega) L(G, \theta, s, \omega) dG(\omega), \tag{8}$$

where $L(G, \theta, s, \omega) = (dP_{\theta,s}/dG)(\omega)$ is the *Radon-Nikodym* derivative of $P_{\theta,s}$ with respect to G. Under appropriate regularity conditions (see [12]), one can differentiate α_t by differentiating inside the integral:

$$\nabla\alpha_t(\theta, s) = \int_\Omega \psi_t(\theta, s, \omega) dG(\omega). \tag{9}$$

where

$$\psi_t(\theta, s, \omega) = L(G, \theta, s, \omega) \nabla_\theta h_t(\theta, s, \omega) + h_t(\theta, s, \omega) \nabla_\theta L(G, \theta, s, \omega). \tag{10}$$

When (9) holds, $\psi_t(\theta, s, \omega)$ can be used to estimate $\nabla\alpha_t(\theta, s)$.

Typically, ω can be viewed as the set of values taken by a finite sequence of independent random variables. For example, let $\omega = (\xi_1, \ldots, \xi_t)$, where for $1 \le i \le t$, ξ_i is the value taken by a continuous random variable (or vector) with density $g_{i,\theta}$. Given X_{i-1}, the value of ξ_i determines the next state X_i (i.e. X_i is a function of (X_{i-1}, ξ_i)). To estimate $\nabla\alpha_t(\theta_n, s)$, an easy choice for G is $P_{\theta_n,s}$. Then, the Radon-Nikodym derivative is the *likelihood ratio*

$$L(P_{\theta_n,s}, \theta, s, \omega) = \prod_{i=1}^{t} \frac{g_{i,\theta}(\xi_i)}{g_{i,\theta_n}(\xi_i)} \tag{11}$$

and its gradient is the *score function*:

$$S(\theta, s, \omega) = \sum_{i=1}^{t} \nabla_\theta \ln g_{i,\theta}(\xi_i). \tag{12}$$

A major problem is that the variance of $S(\theta, s, \omega)$ (and of the LR gradient estimator) typically increases linearly with t. In practice, there is a tradeoff between bias and variance: t_n must be increased with n, but not too fast.

When the system possesses a readily identifiable regenerative structure, α can be written as the quotient of two functions, and a LR approach can be used to obtain an estimator of the derivative of the quotient, for each component of θ. See [5, 7, 17] for more details. There is still a bias problem and t_n must still go to infinity, because that approach involves the expectation of a ratio, but the variance now decreases linearly instead of increasing with the simulation length. That approach could be practical if the regenerative cycles are not too long.

2.4. Infinitesimal Perturbation analysis (IPA)

Here, we define the sample space in such a way that $P_{\theta,s}$ is independent of θ. For instance, one can view ω as a sequence of independent $U(0,1)$ variates. Then, $L(P_{\theta_n,s_n}, \theta, s, \omega) = 1$ and (10) becomes:

$$\psi_t(\theta, s, \omega) = \nabla_\theta h_t(\theta, s, \omega). \tag{13}$$

This is the usual IPA gradient estimator for $\nabla\alpha_t(\theta, s)$ [8, 9, 21, 22].

3. A GI/G/1 QUEUE

Consider a GI/G/1 queue [2] with interarrival and service-time distributions A and B_θ respectively, both with finite expectations and variances. The latter depends on a parameter $\theta \in \bar{\Theta} = [\ell_1, \ell_2] \subset \mathbb{R}$ and has a corresponding density function b_θ. We assume that for all $\theta \in \bar{\Theta}$, the system is stable. Let $w(\theta)$ be the average sojourn time in the system per customer, in steady-state, at parameter level θ. The objective function is defined by

$$\alpha(\theta) = w(\theta) + C(\theta). \tag{14}$$

where $C : \bar{\Theta} \mapsto \mathbb{R}$. We want to minimize $\alpha(\theta)$ (assumed strictly unimodal) over $\Theta = [a, b]$, where $\ell_1 < a < b < \ell_2$. For many distributions, $\alpha(\theta)$ and its minimizer θ^* can be computed analytically or numerically. But let us ignore this momentarily and try to solve the problem using SA, combined with different gradient estimation methods. The solutions of numerical examples can then be compared to the true optimal solutions for an empirical evaluation.

A GI/G/1 queue can be described in terms of a discrete-time Markov chain via Lindley's equation. Let W_i, ζ_i, and $X_i = W_i + \zeta_i$ be the *waiting* time, *service* time, and *system* time for the i-th customer, and ν_i be the time between arrivals of the $(i-1)$-th and i-th customer (for $i = 1$, it is the time until the first arrival). For our purposes, the system time X_i will be the state of the chain at step i. The state space is $S = [0, \infty)$ and $X_0 = s$ is the initial state. $X_0 = 0$ corresponds to an initially empty system. For $i \ge 0$, one has

$$X_i := (X_{i-1} - \nu_i)^+ + \zeta_i \tag{15}$$

where x^+ means $\max(x, 0)$. Since $C(\theta)$ is deterministic, we will estimate only the derivative of $w(\theta)$ and then add $C'(\theta)$ separately to Y_n. Therefore, here, the notation differs slightly from that of the previous sections: $f(\theta, X_i) = X_i$ and $h_t(\theta, s, \omega)$ represents the *average* system time for the t customers who leave during a given subrun of length t (customers). Let $\tilde{S} = [0, c]$ for some (perhaps large) constant c. Let $w_t(\theta, s) = E_{\theta,s}[h_t(\theta, s, \omega)]$ and $\alpha_t(\theta, s) = w_t(\theta, s) + C(\theta)$. We assume that α is continuously differentiable and strictly unimodal in $\bar{\Theta}$. We also need the following assumptions.

ASSUMPTION 1.

(i) The set $\{\zeta \geq 0 \mid b_\theta(\zeta) > 0\}$, *which is the support of* b_θ, *is independent of* θ. *Call it* Δ.

(ii) Everywhere in $\bar{\Theta}$, $b_\theta(\zeta)$ *is continuously differentiable with respect to* θ, *for each* $\zeta \geq 0$, *and continuous in* ζ.

(iii) For each θ *in* $\bar{\Theta}$, b_θ *has a finite Laplace transform in a neighborhood of zero.*

(iv) For each θ_0 *in* $\bar{\Theta}$, $\lim_{\theta \to \theta_0} \sup_{\zeta \in \Delta} b_\theta(\zeta)/b_{\theta_0}(\zeta) = 1$.

(v) For each θ_0 *in* $\bar{\Theta}$, *there exists* $\Psi_{\theta_0} : (0,\infty) \to \mathbb{R}$ *and a neighborhood* Υ_{θ_0} *of* θ_0 *such that* $\sup_{\theta \in \Upsilon_{\theta_0}} |\partial b_\theta(\zeta)/\partial\theta|/b_{\theta_0}(\zeta) \leq \Psi_{\theta_0}(\zeta)$ *for all* ζ, *and* $\sup_{\theta_0 \in \bar{\Theta}} E_{\theta_0}[\Psi^4_{\theta_0}(\zeta)] < \infty$. ∎

ASSUMPTION 2.

(i) Suppose that ζ_j *is generated by inversion* [3]: $\zeta_j = B_\theta^{-1}(U_j) \stackrel{\text{def}}{=} \inf\{\zeta \mid B_\theta(\zeta) \geq U_j\}$, *where* U_j *is a* $U(0,1)$ *variate.*

(ii) There is a distribution $\tilde{B}$ *such that* $\sup_{\theta \in \bar{\Theta}} B_\theta^{-1}(u) \leq \tilde{B}^{-1}(u)$ *for each* u. *The queue remains stable when the service times are generated according to* $\tilde{B}$. *Also,* $\int_0^1 (\tilde{B}^{-1}(u))^4 du < \infty$ *(finite second moment).*

(iii) $B_\theta^{-1}(u)$ *is differentiable in* θ *for each* $u \in (0,1)$.

(iv) There exists a measurable function $\Gamma : (0,1) \mapsto \mathbb{R}$ *such that* $\sup_{\theta \in \bar{\Theta}} |\partial B_\theta^{-1}(u)/\partial\theta| \leq \Gamma(u)$ *for each* $u \in (0,1)$ *and such that* $\int_0^1 (\Gamma(u))^4 du < \infty$. ∎

In [14], we prove that under Assumption 1, each $\alpha_t(\cdot, s)$ is continuously differentiable, and that under Assumption 2 and a mild additional condition, the uniform convergence conditions (3–4) hold. Note that some of these assumptions can be relaxed, but at the cost of getting more complicated.

In the context of this example, t_n represents the number of customers for each subrun of iteration n, except for the regenerative methods, where it represents the number of regenerative cycles per subrun. Regeneration points occur at the beginning of busy periods, i.e. when $X_i = s_0 = 0$. The following propositions are proven in [14].

PROPOSITION 2. *Consider SA with FD, with* $t_n \to \infty$, $c_n \to 0$, $\sum_{n=1}^\infty t_n^{-1}(nc_n)^{-2} < \infty$, *and* $s_n^- = s_n^+$ *for all* n. *Suppose Assumptions 1 and 2 and (3–4) hold. Then,* $\theta_n \to \theta^*$ *with probability one.*

For LR, one can view ω as representing the set of all interarrival and service times generated during a given subrun. One gets the score function (12) with $\xi_i = (\nu_i, \zeta_i)$ and, since only the service times depend on θ, one obtains:

$$Y_n = C'(\theta_n) + h_t(\theta, s, \omega) \sum_{i=1}^{t} \frac{\partial}{\partial\theta} \ln b_\theta(\zeta_i). \tag{16}$$

For the regenerative case, one has

$$Y_n = C'(\theta_n) + \left(\sum_{j=1}^{t_n} \tau_j \sum_{j=1}^{t_n} h_j S_j - \sum_{j=1}^{t_n} h_j \sum_{j=1}^{t_n} \tau_j S_j \right) \left(\sum_{j=1}^{t_n} \tau_j \right)^{-2} \tag{17}$$

where for $j = 1, \ldots, t_n$, τ_j is the number of departures during the j-th regenerative cycle, h_j is the total system time for those τ_j customers who left during that cycle, and $S_j = \sum_{i=1}^{\tau_j} \partial \ln b_\theta(\zeta_i)/\partial\theta$.

PROPOSITION 3. *Suppose that Assumption 1 and (3–4) hold, that* $\sup_{\theta \in \bar{\Theta}} E_\theta[\zeta^8 + \Psi_\theta^8(\zeta)] < \infty$, *and that one uses SA with LR as described above. With the truncated horizon approach, if* $s_n \in \tilde{S}$ *for all* n, $t_n \to \infty$, *and* $\sum_{n=1}^\infty t_n n^{-2} < \infty$, *then* $\theta_n \to \theta^*$ *with probability one. With the regenerative approach, if* $t_n \to \infty$, *then* $\theta_n \to \theta^*$ *with probability one.*

For fixed t, one can *decompose* the cost and estimate the gradient of the waiting time of each individual customer separately. The score function $S_i^{(i)}(\theta, s, \omega)$ associated to a given customer i need not be the sum of all t terms as above, but could include only the terms that correspond to the ζ_j's which can influence that customer's system time. It is

$$S_i^{(i)}(\theta, s, \omega) = \sum_{j=1}^{i} \frac{\partial}{\partial\theta} \ln b_\theta(\zeta_j)$$

and the estimator of $\nabla\alpha_t(\theta, s)$ becomes

$$C'(\theta) + \frac{1}{t} \sum_{i=1}^{t} X_i \sum_{j=1}^{i} \frac{\partial}{\partial\theta} \ln b_\theta(\zeta_j). \tag{18}$$

This LR estimator has approximately half the number of terms as the previous one. Another way of reducing variance is to use the estimator $C'(\theta) + (1/t)\sum_{i=1}^t (X_i - w(\theta)) S_i^{(i)}(\theta, s, \omega)$, in which $w(\theta)$, when unknown, can be replaced by an independent estimate. See [13] for further details.

For IPA [12, 21, 23], the idea is to differentiate (15) for a fixed set of underlying uniform variates. An infinitesimal perturbation on ζ_j affects the system time of customer j and of all the customers (if any) following him in the same busy period. Therefore,

$$\nabla_\theta h_t(\theta, s, \omega) = \frac{1}{t} \sum_{i=1}^{t} \sum_{j \in \Xi_i} \frac{\partial B_\theta^{-1}(U_j)}{\partial\theta} \tag{19}$$

where Ξ_i is the set containing customer i and all the customers that precede him in the same busy period (if any). We call the inside sum in (19) the *IPA accumulator*. When the state is not reset to s_0 between iterations, we can consider two variants of this: one in which Ξ_i can contain only customers who left during the current iteration (the IPA accumulator is reset to zero between iterations) and one in which it can contain customers from the previous iterations (which have indices $j \leq 0$ in (19)). In the latter case, the value a_n of the IPA accumulator at the beginning of iteration n must be stored and could be viewed as part of s_n. For IPA with a regenerative approach, the estimator is defined as in (19), but with t denoting the number of customers that leave during the t_n regenerative cycles. In that case, t is the value taken by a random variable T_n.

PROPOSITION 4. *Suppose that (3–4) hold and that one uses SA with IPA, under Assumptions 1 and 2, with* $s_1 = s_0$ *and with* $t_n \to \infty$. *Then, both with the truncated horizon approach with* $s_n \in \tilde{S}$ *and* $a_n = 0$ *for all* n, *and with the regenerative approach,* $\theta_n \to \theta^*$ *with probability one.*

If the IPA accumulator a_n is not reset to 0 between iterations, IPA has the stronger property, for this particular example and under mild additional conditions, that even when using a truncated horizon t_n that is constant with n, SA converges *weakly* to the optimizer. A proof is given in [14], based on a theorem of Kushner and Shwartz [11]. In the regenerative case, SA does not converge to the optimum in general if t_n does not converge to infinity.

4. Numerical experiments with an M/M/1 queue

Consider an M/M/1 queue with arrival rate $\lambda = 1$ and mean service time θ. One has $B_\theta(\zeta) = 1 - e^{-\zeta/\theta}$. Let $\Theta = [.01, .95]$ and $C(\theta) = 1/\theta$. Then, $w(\theta) = \theta/(1-\theta)$ and $\theta^* = 0.5$. Assumptions 1 and 2 are easily verified.

We performed the following experiments. For each *variant*, i.e. each way of combining SA with a specific derivative estimator, we made 10 simulation runs, each yielding an estimation of θ^*. The 10 initial parameter values were randomly chosen, uniformly over $[.01, .95]$, and the initial state was s_0 (an empty system). Across the algorithms, we used common random numbers and the same set of initial parameter values. This means that the different entries of Table 1 are strongly correlated. Each run was stopped after 10^6 ends of service. The final state of each simulation subrun was taken as the initial state for the next one, except when stated otherwise. For FDC, the initial state s_{n+1} was the final state of the subrun at iteration n with parameter value the closest to θ_{n+1}.

For each variant, we computed the empirical mean $\bar{\theta}$, standard deviation s_d and standard error s_e of the N retained parameter values. If y_i denotes the retained parameter value for run i (i.e. the last value of θ_n), the above quantities are defined by

$$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N} y_i; \qquad s_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{\theta})^2; \qquad s_e^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \theta^*)^2. \tag{20}$$

We also computed 95% confidence intervals I_θ on the expectation of $\bar{\theta}$, assuming that $\sqrt{N}(\bar{\theta} - E(\bar{\theta}))/s_d$ follows a Student distribution with $N-1$ degrees of freedom. The results appear in the third column of Table 1.

In the Table, LRR refers to the regenerative version of LR given in (17), while IPAR refers to the regenerative version of IPA. The symbol -0 means that the state was reset to $s_0 = 0$ at the beginning of each iteration. The symbol -Z following IPA means that the IPA accumulator was reset to 0 between iterations. LR-D means the "decomposed" version of LR given by (18). LR-C [LR-DC] means LR [LR-D] in which $h_t(\theta, s, \omega)$ was replaced by $h_t(\theta, s, \omega) - 1$. This does not change the expectation of $\psi_t(\theta, s, \omega)$, but reduces its variance from $O(t)$ to $O(1)$ at $\theta = \theta^*$, because $w(\theta^*) = 1$ (see [13]). In all cases, we had $\gamma_n = 1/n$. We took $c_n = 0.1n^{-1/6}$ for FD and $c_n = 0.1n^{-1/5}$ for FDC. For FDC, we also tried $c_n = 0.001n^{-2}$, which is denoted by FDC-NN.

We see that IPA performs well, even when t_n is fixed at a small constant. IPA-Z, IPAR, FDC, and FDC-NN, with a linearly increasing t_n, are approximately as good. When t_n is fixed to a small constant, convergence is also quick with FDC, IPA-Z, or IPAR (small s_d), but the standard error s_e is very large, which indicates that convergence is not to the right value. Even for $t_n = 100$, the bias is still quite apparent for FDC. The LR methods in general have trouble due to their large associated variance, and large bias when t_n grows slowly. LR with $t_n = n^p$ has large variance for large p, and for small p, the bias goes down much too slowly compared to the variance. As a result, the confidence interval I_θ, based on the N final values of θ_n, is very likely not to cover θ^*. This is what happens, for instance, with $p = 1/3$. Among the truncated-horizon variants, LR-C and LR-CD provide significant improvements over LR. The LR variant that gives the best results here is LRR (regenerative) with t_n increasing linearly. With $t_n = n^{1/2}$, both LRR and FDC have the same bias problem as described above: the bias goes down too slowly and I_θ does not contain θ^*. Nevertheless, they converge (slowly) to the right answer (we verified it empirically with longer simulation runs).

Independent sets of experiments were also performed with $\bar{T} = 10^5$ and the results were quite similar to the ones given here [4]. We also made experiments with $C(\theta) = 1/(25\theta)$ (for which $\theta^* = 1/6$) and $C(\theta) = 25\theta$ (for which $\theta^* = 5/6$). The results appear in the last two columns of Table 1.

		$C_1 = 1$ $(\theta^* = 1/2)$		$C_1 = 1/25$ $(\theta^* = 1/6)$		$C_1 = 25$ $(\theta^* = 5/6)$	
	T_n	s_d	s_e	s_d	s_e	s_d	s_e
FD	n	.00979	.00967				
FD	$100+n$	.01075	.01044				
FDC	5	.00149	.15343 ◁				
FDC	100	.00340	.00721 ◁				
FDC	n	.00193	.00184	.00030	.00030	.02354	.02234
FDC	$100+n$	.00204	.00198	.00027	.00029	.02875	.02824
FDC-0	n	.00243	.00231	.00039	.00037	.03019	.02867
FDC-NN	n	.00203	.00196	.00031	.00031	.02270	.02177
FDC	$n^{1/2}$	.00181	.00684 ◁				
IPA	1	.00227	.00217				
IPA	10	.00227	.00216	.00053	.00051	.02402	.02575
IPA	100	.00229	.00219				
IPA	n	.00195	.00185	.00046	.00044	.03208	.03416
IPA	$100+n$	.00203	.00193	.00046	.00043	.02685	.02849
IPA-Z	10	.00169	.07365 ◁				
IPA-Z	n	.00192	.00189	.00046	.00044	.02449	.02597
IPA-0	n	.00246	.00233	.00042	.00040	.01721	.01956
IPAR	5	.00228	.06175 ◁				
IPAR	n	.00200	.00197	.00046	.00044	.02981	.03110
LR	$n^{1/3}$	.01221	.02062 ◁				
LR	$n^{1/2}$	.03012	.02876	.02454	.02355	.04473	.05214
LR	$n^{2/3}$	.07494	.07115				
LR-C	$n^{1/2}$	.00772	.00749	.00221	.00291 ◁	.03433	.04864 ◁
LR-C0	$n^{1/2}$	.00709	.00725				
LR-D	$n^{1/2}$	.01502	.01658				
LR-CD	$n^{1/2}$	.00533	.00615	.00175	.00176	.03000	.05141 ◁
LR-CD	$n^{2/3}$	.00706	.00688	.00264	.00255	.04893	.04857
LRR	n	.00447	.00453	.00124	.00118	.07608	.07446
LRR	$n^{1/2}$	.00443	.01775 ◁				

Table 1: Some experimental results for 10 times 10^6 customers.
For the values marked with ◁, the 95% confidence interval does not contain θ^*.

ACKNOWLEDGMENTS

The work of the first author was supported by NSERC-Canada grant # A5463 and FCAR-Québec grant # EQ2831. The third author's work was supported by the IBM corporation under SUR-SST contract 12480042 and by the U.S. Army Research Office under contract DAAL-03-88-K-0063.

REFERENCES

[1] Aleksandrov, V. M., V. I. Sysoyev and V. V. Shemeneva, "Stochastic Optimization", *Engineering Cybernetics*, 5 (1968), 11–16.

[2] Asmussen, S., *Applied Probability and Queues*, Wiley, 1987.

[3] Bratley, P., B. L. Fox, and L. E. Schrage, *A Guide to Simulation*, Springer-Verlag, New York, Second Edition, 1987.

[4] Giroux, N. "Optimisation Stochastique de Type Monte Carlo", Mémoire de maîtrise, dépt. d'informatique, Univ. Laval, jan. 1989.

[5] Glynn, P. W. "Likelihood Ratio Gradient Estimation: an Overview", *Proceedings of the Winter Simulation Conference 1987*, IEEE Press (1987), 366–375.

[6] Glynn, P. W. "Optimization of Stochastic Systems Via Simulation", *Proceedings of the Winter Simulation Conference 1989*, IEEE Press (1989), 90–105.

[7] Glynn, P. W. "Likelihood Ratio Gradient Estimation for Stochastic Systems", *Communications of the ACM*, **33**, 10 (1990), 75–84.

[8] Heidelberger, P., Cao, X.-R., Zazanis, M. A. and Suri, R., "Convergence Properties of Infinitesimal Perturbation Analysis Estimates", *Management Science*, **34**, 11 (1989), 1281–1302.

[9] Ho, Y.-C., "Performance Evaluation and Perturbation Analysis of Discrete Event Dynamic Systems", *IEEE Transactions of Automatic Control*, **AC-32**, 7 (1987), 563–572.

[10] Kushner, H. J. and Clark, D. S., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, Applied Math. Sciences, vol. 26, 1978.

[11] Kushner, H. J. and Shwartz, A., "An Invariant Measure Approach to the Convergence of Stochastic Approximations with State Dependent Noise", *SIAM J. on Control and Optim.*, **22**, 1 (1984), 13–24.

[12] L'Ecuyer, P., "A Unified View of the IPA, SF, and LR Gradient Estimation Techniques", *Management Science*, **36**, 11 (1990), 1364–1383.

[13] L'Ecuyer, P. and Glynn, P. W., "A Control Variate Scheme for Likelihood Ratio Gradient Estimation", In preparation (1990).

[14] L'Ecuyer, P., Giroux, N., and Glynn, P. W., "Stochastic Optimization by Simulation: Convergence Proofs and Experimental Results for the GI/G/1 Queue", manuscript, 1990.

[15] Meketon, M. S., "Optimization in Simulation: a Survey of Recent Results", *Proceedings of the Winter Simulation Conference 1987*, IEEE Press (1987), 58–67.

[16] Pflug, G. Ch., "On-line Optimization of Simulated Markovian Processes", *Math. of Oper. Res.*, **15**, 3 (1990), 381–395.

[17] Reiman, M. I. and Weiss, A., "Sensitivity Analysis for Simulation via Likelihood Ratios", *Operations Research*, **37**, 5 (1989), 830–844.

[18] Rubinstein, R. Y., *Monte-Carlo Optimization, Simulation and Sensitivity of Queueing Networks*, Wiley, 1986.

[19] Rubinstein, R. Y., "The Score Function Approach for Sensitivity Analysis of Computer Simulation Models", *Math. and Computers in Simulation*, **28** (1986), 351–379.

[20] Rubinstein, R. Y., "Sensitivity Analysis and Performance Extrapolation for Computer Simulation Models", *Operations Research*, **37**, 1 (1989), 72–81.

[21] Suri, R., "Infinitesimal Perturbation Analysis of General Discrete Event Dynamic Systems", *J. of the ACM*, **34**, 3 (1987), 686–717.

[22] Suri, R., "Perturbation Analysis: The State of the Art and Research Issues Explained via the GI/G/1 Queue", *Proceedings of the IEEE*, **77**, 1 (1989), 114–137.

[23] Suri, R. and Leung, Y. T., "Single Run Optimization of Discrete Event Simulations—An Empirical Study Using the M/M/1 Queue", *IIE Transactions*, **21**, 1 (1989), 35–49.