When, and When Not to Use Warm-up Periods in Discrete Event Simulation

Winfried K. Grassmann
Department of Computer Science, University of Saskatchewan
110 Science Court, Saskatoon, SK S7N 5C9
grassman@cs.usask.ca

ABSTRACT

This paper demonstrates, both theoretically and by using numerical examples, that if one has good starting states, one should not use warmup periods in discrete event simulation when estimating equilibrium expectations by means of time averages. The numerical methods used are deterministic, and they are based on randomization or uniformization. We also show that if estimating expectations of sums by simulation, good starting states are sometimes difficult to find and are often inconvenient, which justifies warmup periods.

Categories and Subject Descriptors

I.5 [**Simulation**]: Output Analysis—*Initialization bias*

General Terms

Initialization, Estimation

Keywords

Discrete Event Simulation, Time Averages, Initialization Bias, Queues, Agent-based Systems

1. INTRODUCTION

A typical objective of a non-terminating discrete event simulation is to find the equilibrium expectation of some quantity $X(t)$. To do this, one typically uses time averages, that is, one forms

$$\bar{X}(T) = \frac{1}{T} \int_0^T X(t)dt. \qquad (1)$$

To start the simulation, one has to choose some initial state, but this will result in an *initialization bias*. Of course, besides the initialization bias, one also has to consider the variance of the time average. To deal with both, one typically uses the mean squared error or MSE (see e.g [1], [19], [20]), which combines the effect of the variance and the bias. Also,

a small MSE insures that the estimator $\bar{X}(T)$ given by (1) is close to equilibrium expectation $E(X)$ to be estimated.

To reduce the initialization bias, and hopefully also the MSE, one can either choose an appropriate initial state, or one can start the simulation at a time $-t_0$, but start collecting data only after time 0. The period from $-t_0$ to 0 is typically called *warmup period* [10] or *initialization phase* [2]. Removing the data collected in the warmup period is often called *transient deletion* [1], [12], [13]. In this paper, we present evidence, both experimentally and theoretically, that warmup periods should not be used, provided one starts in a frequently visited state. Both, the experiments and the theoretical discussions are conducted by using Markov chains. We note that means and variances of time averages arising within Markov chains can be calculated by deterministic numerical methods, that is, non-simulation methods.

There is a fair amount of literature discussing the initialization bias. Indeed, most textbooks cover this topic (see e.g. [2], [3] and [10]). A survey of the initialization bias problem is contained in [13]. Experimental results can be found in [12], and some credibility issues are discussed in [14].

Pawlikowski [13] discusses a total of eleven methods to deal with the initialization bias. Nine of these methods use statistical tests based on sample functions to determine when the equilibrium is reached. Though such tests are certainly useful, our contribution has a different flavor. We are looking at general rules to decide how to deal with the initialization bias. Here is an example of what we mean: Awad and Glynn [1] (see also [4]) suggest to first estimate the equilibrium probabilities, and to select the starting state nondeterministically, but according to these equilibrium probabilities. This obviously removes any bias. In this paper, we show that this method does not give the lowest MSE, a somewhat counterintuitive result. In fact, we show that there is always a fixed state that will give a lower MSE than the one selected based on the equilibrium probabilities, provided the simulation time is long enough. Hence, the effort should be geared toward finding a good starting state rather than using a stochastic initial state based on the equilibrium probabilities.

As was pointed out in [7], the equilibrium distributions, being distributions, must be distinguished from the equilibria of deterministic systems. However, there are sample functions that are almost deterministic. In particular, if the sample function $X(t)$ of interest is the sum of uncorrelated or weakly correlated random variables, limit theorems apply, and the stochastic component of $X(t)$ is reduced. How does

this affect the initialization bias will also be investigated in this paper.

The outline of this paper is as follows: In Section 2, we provide some background on Markov modeling, which is basic to our further discussion. In Section 3, we present some examples that demonstrate that selecting the initial state according to the equilibrium probabilities does not lead to the smallest MSE, and we support this result by mathematical arguments. Section 4 deals with the case where the statistic of interest is a sum, and how to determine the warmup period if a good starting state is not available. Section 5 will provide some conclusions.

2. THE MARKOV MODELING APPROACH

Our investigations are based on Markov chains. This does not mean that the variable which is observed in the simulation is necessarily Markovian. Indeed, by adding so-called supplementary variables, one can convert practically any discrete-event system into a Markov chain as shown in [7]: one can use phase-type distributions as supplementary variables, and phase-type distributions are dense on the positive axis, which means that any distribution can be approximated arbitrarily closely by a phase-type distribution. If one has supplementary variables, one has to distinguish between the variable of interest, say $X(t)$, and the state of the system, which is a vector that includes all state variables, including the supplementary variables. Hence, besides $X(t)$, the variable of interest, we have to keep track of the state at time t, say $M(t)$, which is typically a vector. The set of all states will be denoted by S, and the set all values the variable X can assume by S_X.

The MSE is influenced by the initial state, that is, by $M(0)$ in the case where there is no warmup period. If a warmup period starts at time $-t_0$, then $M(0)$ becomes a random vector. The probability that $M(0) = i$, $i \in S$ will be denoted by $\pi_i(0)$. We assume that the process has a unique (multidimensional) equilibrium distribution which will be denoted by π_i. If the warmup period t_0 approaches infinity, $\pi_i(0)$ approaches π_i. Hence, as pointed out in [1] and [4], choosing the state at time 0 with probability π_i corresponds to a warmup period of infinite length. Clearly, if starting in this sense with the equilibrium probabilities does not reduce the MSE, then warmup periods should not be used. This will be a salient point in the discussion that follows.

As mentioned, we use deterministic numerical methods for our investigations. Deterministic numerical methods have also been used by others such as [11] and [19]: these authors used explicit formulas available for the M/M/1 queue. Our algorithms, on the other hand, can be deal with arbitrary Markov chains. This allows us to deal with a great variety of models, as will be shown. In contrast, [19] had to use simulation already for the M/M/c queue, a system that can easily be analysed by our algorithms.

Deterministic algorithms tend to be more accurate than simulation, and for small to medium sized models, they are much faster. Of course, if the solution of a large problem is required, most deterministic numerical methods are subject to the curse of dimensionality, which implies that the computational effort increases exponentially with the number of state variables. Hence, as state variables are added, the size of the problem increases rapidly. In case of Markov chains, different methods become competitive as the size of the problem increases. For instance, to find steady-state solutions in small Markov chains, one can use direct methods, such as Gaussian elimination or the GTH method [17]. For large Markov chains, iterative methods are faster. From a certain point onwards, simulation methods become competitive, and as the number of states increases further, they are the only practical method. For demonstration purposes, however, it is better to use small examples: it is much easier to highlight the essential features with small and simple examples than with the large examples one encounters in industrial applications. Hence, for our purposes, deterministic methods are preferable. Indeed, all our examples were done in Excel, using VBA, a rather slow interpreter. The models were run of an IBM Laptop.

It is convenient to express the Markovian systems using events: in Markov modeling, events occur completely at random, and they change the state variables in a predictable way. We assume that there are s events. The event-based formalism is then converted to a continuous-time Markov chain with transition matrix $A = [a_{ij}]$. Typically, A is sparse. If each of the s events can occur in every state, and there are N states, then A is an $N \times N$ matrix with sN non-zero entries, Typically, some events cannot occur in certain states, say departures cannot occur from an empty queueing system. Therefore, A has typically fewer than sN non-zero entries, and since s tends to be small compared to N, A becomes sparse.

The method used here to find the MSE is based on randomization. Randomization involves the conversion of the transition matrix A into the stochastic matrix $P = A/q + I$, where q is chosen such that all entries of P are positive, which is true if q equals or exceeds the largest leaving rate $-a_{ii}$, and I is the identity matrix. The Markov chain with transition matrix P is then randomized, that is, instead of using constant intervals between steps, one uses exponential random variables with rate q. For details of this method, see e.g. [17]. This method is extremely efficient for finding transient solutions of continuous-time Markov chains, and it is implemented in many packages, such as [16]. To find all probabilities $\pi_i(t)$ to be in state i at time t with a precision better than α by randomization, one needs roughly $sN^2(qt + z_\alpha \sqrt{qt})$ multiplications and additions, where z_α can be found from normal tables. The z_α arises because the Poisson distribution approaches the normal distribution as qt increases. We note that the time and space complexity of the randomization method is comparable to that of finding equilibrium probabilities.

Here, we need means and variances of time averages, and to obtain these, the randomization method can be modified to accomplish this. Actually, as shown in [6], to find the variance of a time average is even slightly faster than the randomization method described above, and the same is true in order to find the the expectation of a time average.

3. SOME COUNTERINTUITIVE RESULTS

The MSE is influenced by both the bias and the variance of $\bar{X}(T)$, where $\bar{X}(T)$ is defined by (1). Here, the bias is defined as

$$B(\bar{X}(T)) = \mathrm{E}(\bar{X}(T)) - \mathrm{E}(X),$$

where $\mathrm{E}(X)$ is the equilibrium expectation to be determined.

Table 1: Measures for the time average of the number in an M/M/1/10 system

$\lambda = 0.8$, $\mu = 1$, $T = 1000$
$E(X) = 2.966$, $Var(X) = 7.561$

$X(0)$	$B(\bar{X}(T))$	$Var(\bar{X}(T))$	MSE
0	-0.0273	0.1703	0.17100
1	-0.0236	0.1705	0.17101
2	-0.0165	0.1709	0.17113
3	-0.0064	0.1714	0.17147
4	0.0061	0.1721	0.17210
5	0.0205	0.1727	0.17309
equil	0.0000	0.1725	0.17247

The mean squared error can be written as

$$
\begin{aligned}
\text{MSE}(\bar{X}(T)) &= E((\bar{X}(T) - E(X))^2) \\
&= Var(\bar{X}(T)) + B^2(\bar{X}(T)). \quad (2)
\end{aligned}
$$

Hence, the MSE is equally influenced by the bias and by $Var(\bar{X}(T))$.

To see how the MSE is influenced by the initial state $M(0)$, consider the M/M/1/10 queue, that is, a one-server queue with Poisson arrivals, exponential service times, and a waiting room of 10. Table 1 contains the results for the time average between 0 and $T = 1000$ for an arrival rate $\lambda = 0.8$ and a service rate $\mu = 1$, resulting in a traffic intensity $\rho = \lambda/\mu = 0.8$. The table shows the bias, the variance and the MSE for the queue starting with $M(0) = X(0)$ customers initially. The lowest MSE is reached when $X(0)$ is set to 0, a result proven in [11]. Note that the expected value is $E(X) = 2.966$, but starting with 3 customers in the system leads to a higher MSE. The last row shows what happens if the initial state is selected according to the equilibrium probabilities of the process as suggested by [1]. In this case, the bias is obviously 0. However, the MSE is higher than when starting with $X(0) \leq 4$, that is, any effort spent for warmup is not only worth nothing, it is even detrimental. To me, this result seemed extremely counterintuitive, and I therefore used a different method to find the MSE as a check, namely one based on eigenvalues suggested by Reynolds [15]. The result matched up to 6 digits [7]. Hence, the effect is real. Returning again to Table 1, one notices, however, that the starting state is really not all that important. The same can be said for queues with $c > 1$ servers, provided $c \leq 10$ [7]. We found that this is also true for Erlang queues. In other words, if one has a reasonable initial state $M(0)$, the initial bias problem is a non-issue, and there should not be any warmup period.

We also ran the M/M/1 queue with finite buffers for other buffer sizes, and other values of ρ, but the results were essentially the same. Also note that if the system starts with the equilibrium probabilities, one can use the variance to find a 95% confidence interval, and this is $2.966 \pm 1.96\sqrt{0.17247} = 2.966 \pm 0.814$. Obviously, setting T to a value below 1000 would increase this confidence interval, and the practical value of a simulation with such a large confidence interval is questionable.

To gain a further understanding of the problem, note that $TB(\bar{X}(T)) = B + o(1/T)$, and that $TVar(\bar{X}(T)) = V + o(1/T)$, where B and V are constants, that is, they are independent of T [8], [18]. As shown in [7], the convergence

is usually quite fast. Hence, we conclude from (2):

$$
\text{MSE}(\bar{X}(T)) = \frac{V}{T} + \frac{B^2}{T^2} + o(1/T). \quad (3)
$$

Consequently, as T increases, $\frac{B^2}{T}$ decreases much faster than $\frac{V}{T}$, and for large enough T, the bias therefore becomes negligible when compared to the variance. For a related result, see [1].

In addition to this, the following theorem is proven in [7].

THEOREM 1. *Let $Var_i(\bar{X}(T))$ be the value of $Var(\bar{X}(T))$, given the system starts in state i. Then, for any stochastic system, $\min_i Var_i(\bar{X}(T)) < Var_{\pi(0)}(\bar{X}(T))$, where $Var_{\pi(0)}(\bar{X}(T))$ is the variance when starting with probability $\pi_i(0)$ in state $i \in S$. Here, we assume that no $\pi_i(0) = 1$.*

The proof follows immediately from the law of total variance, as shown in [7]. The theorem implies that there is at least one state i which has a lower $Var_i(\bar{X}(T))$ then the variance obtained by starting with the equilibrium probabilities. Let i^* be such a state. For large enough T, equation (3) indicates that the bias becomes negligible compared to the variance, which means that when setting $M(0)$ to i^*, $\text{MSE}_i(\bar{X}(T)$ is less than $\text{MSE}_\pi(\bar{X}(T))$, and no warmup period should be used. Here, the subscript i denotes the initial state, and the subscript π indicates that the initial state is chosen according to the equilibrium distribution. This convention will be used freely in the remainder of this paper. For small systems, values of i^* are typically not difficult to find. For instance, for M/M/c queues, one can choose the initial number in the system to be approximately equal to the traffic intensity, that is, the arrival rate over the service rate.

Note that Theorem 1 holds for any Markov chain, no matter how large, and it holds also if supplementary variables are present. It even holds when the supplementary variables are continuous. The theorem therefore remains valid for non-Markovian systems, provided that by introducing a sufficient number of supplementary variables, the system can be made Markovian.

Another reason why warmup periods are not always beneficial is the following. If one starts in a state i which is of reasonable importance, say it is visited often, then one is almost guaranteed to visit this state quite often during a simulation that is long enough to yield significant results. In Markov chains, the system regenerates at these points, and one obtains several probabilistically undistinguishable sub-processes, all starting in state i. Because of this, all these subprocesses should all be treated the same way, and they could be interchanged without effecting the underlying the stochastic properties of the process. Hence, why should the first subprocess be singled out and be given a warmup period? On the other hand, it does make sense to have a warmup period if the initial state is an outlier, or if it is a transient state, that is, a state where the expected number of visits is finite.

4. ESTIMATING EXPECTATIONS OF SUMS

In the previous section, we showed that there is a state i^* such that if the system is in i^* at time 0, then the MSE is minimized provided the data collection time is long enough. Our examples, together with the examples given in [7], showed

that in many cases, the state the system is in at time 0 has no major effect on the MSE. We will call the state the system is in at time 0 *time-zero* state. If there is no warmup, that is, if t_0 is zero, then the time-zero state is equal to the starting state. A time-zero state will be called a *good time-zero* state if the MSE corresponding to this state is close to the minimal MSE, that is, close to MSE_{i*}.

One can sometimes find a good time-zero state by inspection: for instance, in the M/M/1 queue, it can be shown that starting with an empty system is best [11]. If good time-zero states cannot be found, a warmup period from time $-t_0$ to time 0 can be used. Warmup periods are also useful if it is inconvenient or impossible find a good time-zero state. For instance, in agent-base systems, one may want to initialize all agents in the same way, using identical initial parameters, and let the simulation determine how the system develops. Often, the length of the warmup period needed to hit a good time-zero state can can be found by considering the structure of the problem. For instance, in a sequential queueing system, one may be interested in the output of the last server. If one starts with an empty system, it will take a while until the first job reaches the last server, and before this happens, the output of the last server is zero, and hence a poor time-zero state. The logical way to mend the situation is to make the warmup period t_0 long enough such that the first job can reach the last server.

Frequently, one has to estimate expectations of sums, and in this case, finding a good initial state can be both difficult and inconvenient. Hence, in this case, warmup periods may be needed. If the sum has only a few terms, it may still be better to have no warmup, but this changes as soon as the number of terms reaches a certain threshold. As we will show, this threshold increases with increasing correlation among the summands.

We consider the threshold in the context of systems containing many components, we will call them *agents*. Let $X(t)$ be a sum, with each agent of the population, or of some well-defined subpopulation contributing a summand. Such systems include many agent-based systems, but also systems such as queueing networks, where one is interested in the number of jobs in a subnetwork, or even the entire network. Agents have attributes, such as age, and a simulation may be used to find the average age, and to do this, one needs the sum of all ages. Properties are attributes that can have only two values, namely 0 and 1: the value 1 indicates that the agent has the property, and 0 that it does not. If Y_i is the value of the property Y for agent i, then the number of agents with property Y in a population of size N is obviously $\sum_{i=1}^{N} Y_i$. In queueing networks, the agents would be the queues, and their attributes the queue lengths.

To find the MSE involving sums, consider first the case where the individual summands $Y_i(t)$ are independent identically distributed random variables. The distributions in question are identical if all agents are initialed in the same way and if their development in time is determined by the same stochastic laws. The distributions are independent if the agents do not influence one another in any way. Though the assumption of independence is not normally true, it will provide useful bounds. Hence, consider the estimator for $\text{E}(X)$ in a population of size N, where $X(t) = \sum_{i=1}^{N} Y_i(t)$. Let $\bar{Y}_i(T)$ be the time average of $Y_i(t)$ in the interval from 0 to T. Since all $\bar{Y}_i(T)$ have the same expectation and variance, we can drop the subscript i, that is $\text{B}(\bar{Y}(T))$ is

the common bias and $\text{Var}(\bar{Y}(T))$ the common variance of the $\bar{Y}_i(t)$. It follows that $\text{Var}(\bar{X}(T)) = N\text{Var}(\bar{Y}(T))$ and $\text{B}(\bar{X}(T)) = N\text{B}(\bar{Y}(T))$. Hence

$$
\begin{aligned}
\text{MSE}(\bar{X}(T)) &= \sum_{i=1}^{N} \text{Var}(\bar{Y}_i(T)) + \left(\sum_{i=1}^{N} \text{B}(\bar{Y}_i^*(T))\right)^2 \\
&= N\text{Var}(\bar{Y}(T)) + N^2\text{B}^2(\bar{Y}(T)). \quad (4)
\end{aligned}
$$

As N increases, the bias gains in weight when compared to the variance. In fact, if N is large, the bias becomes huge, and the asymptotic suggested by (2) becomes problematic. To illustrate this, consider an agent-based system, with all agents initialized the same way. Of course, the equilibrium probability of such an arrangement is small, which means such a state is a poor starting state, and it becomes worse as the population size N increases. Nevertheless, this type of initialization can be very convenient.

After the warmup, which terminates at time 0, the probability to be in state i is $\pi_i(0)$. Of course, the best that one can hope for is the at time 0, $\pi_i(0) = \pi_i$, where the π_i are the equilibrium probabilities, and it thus makes sense to concentrate on this case. Starting with the equilibrium probabilities brings the bias to 0, and $\text{MSE}_\pi(\bar{X}(T)) = \text{Var}_\pi(\bar{X}(T))$. Obviously, one should not start in equilibrium if

$$
\text{Var}_\pi(\bar{X}(T)) > \text{MSE}_i(\bar{X}(T)) = \text{Var}_i(\bar{X}(T)) + \text{B}_i(\bar{X}(T))^2.
$$

If equation (4) holds, this is equivalent to

$$
N\text{Var}_\pi(\bar{Y}(T)) > N\text{Var}_i(\bar{Y}(T)) + N^2\text{B}_i(\bar{Y}(T))^2
$$

or

$$
\frac{\text{Var}_\pi(\bar{Y}(T)) - \text{Var}_i(\bar{Y}(T))}{B_i(\bar{Y}(T))^2} > N. \quad (5)
$$

If N is below the left hand side of this inequality, then it does not make sense at all to have a warmup period. Since warmup periods require computational resources, it is very well possible that even when N is violating (5), warmup periods are not worthwhile. Hence, (5) is, in this sense, only a lower bound for the threshold requiring a warmup.

To show how (5) is applied, we assume that we have a population of N M/M/1/10 queues, all of them identical and independent, and we ask as to when considering a warmup period makes sense. From Table 1, we obtain, if 0 is the starting state

$$
\frac{\text{Var}_\pi(\bar{Y}(T)) - \text{Var}_0(\bar{Y}(T))}{B_0(\bar{Y}(T))}
$$
$$
= (0.17247 - 0.17100)/0.0273^2 = 1.97.
$$

Hence, for $N < 1.97$, there is no point using a warmup period, and for practical purposes, we may use a threshold of 2. Even as N increases beyond that, warmup periods may not be required. The are only needed if N is somewhat larger than this lower bound.

The threshold for the need of warmup periods increases if the terms of a sum are correlated. In the extreme case of perfect positive correlation, the variance also increases with N^2, and the relative importance of the variance relative to the squared bias does not change as N increases. In practical cases, the correlation coefficients are somewhere between 0 and 1, that is, one has neither uncorrelated random variables, nor perfectly correlated random variables. A little reflection shows that in this case, (5) can still serve as an

upper bound. To see how correlation will affect the value of N satisfying (5), we introduce the following example.

Consider a stochastic version of what is called the SIRS model [9], for Susceptible, Infected, Recovered and Susceptible in systems analysis. In this model, there is a population of size N, say a population of humans, where each person is either susceptible to a certain illness, or he/she has the illness, or he/she has recovered and has immunity. After a certain time, immunity is lost. We assume that the rate of a susceptible person to get infected is λ_S, that the rate of an infected person to recover is λ_I, and the rate of a recovered person to lose immunity and become susceptible again is λ_R. We use two models: infections through some external source, such as contaminated water, and infections through contact with other people. In the first model, λ_S is independent of the number of infected people, and the health status of the agents are independent random variables, but in the second model, the corresponding random variables are not independent. Both models are Markovian, and the two models have the same recovery rate $\lambda_I = 10$, and the same rate of losing immunity $\lambda_R = 1$. The two models differ in the infection rates: for the external source model, $\lambda_S = 1$, whereas in the infection through contact model, $\lambda_S = 0.01 + 0.8\sqrt{X_I}/\sqrt{N}$. The simulation runs for a time interval of $T = 100$. Provided one wants to find the expected number of infected people X_S, for which values of N should a warmup period be considered? To decide this, we use (5), with $Y_i(T) = 1$ if person i is infected, and $Y_i(t) = 0$ otherwise. If all agents are independent and initialized in the same way, then it is best to initialize the agents to be neither susceptible nor infected, leading to a values of $\mathrm{Var}(\bar{Y}(T)) = 6.45318\text{E-}05$ and $\mathrm{B}(\bar{Y}(T)) = -0.000272109$. The value of $\mathrm{Var}_\pi(\bar{Y}(T))$ is then 6.48495E-05, and the bound given by (5) becomes 4.29. Hence, there is no point to have a warmup period before the population reaches a size of 5.

At which population size does it pay to have a warmup period in the case of infection through contact? One could use (5), but this formula requires values for $\mathrm{Var}(\bar{Y}(T))$ and $\mathrm{B}(\bar{Y}(T))$. Somewhat arbitrarily, we used the values obtained while $N = 1$. If the initialization must be the same for all agents, then the best starting state is again $X_S = X_I = 0$, and the values for $\mathrm{Var}(\bar{Y}(T)) = 0.000017357$, $\mathrm{B}(\bar{Y}(T)) = -0.000097251$ and $\mathrm{Var}(\bar{Y}(T)) = 0.000017524$, respectively. Hence, (5) yields 17.5.

Instead of using this approximation, we calculated the exact values for MSE_{i_0} and MSE_π, where i_0 is the state with $X_S = X_I = 0$. The results are given in Table 4. Warmup periods are obviously only beneficial if $\mathrm{MSE}_\pi < \mathrm{MSE}_{i_0}$, and according to Table 4, this happens only for $N > 14$. However, even for N=20, the difference between MSE_{i_0} and MSE_π is hardly noticeable, and a warmup period, which requires computational resources, may not be advantageous. Of course, as the population increases, the difference between MSE_π and MSE_{i_0} becomes larger, and at some point, one has to find either a good starting state, or one has to use a warmup period. In the latter case, the statistical methods reviewed in [13] for determining the length of the warmup seem to be appropriate.

To summarize, our example indicates that when estimating expectations of sums with only few terms, warmup periods are often detrimental. However, as the number of terms in the sum increases, warmup periods become advantageous. The crossover point between no warmup period

Table 2: Effect of warmup in infection through contact model

	MSE_{i_0}	MSE_π
1	1.73674E-05	1.75238E-05
2	3.3156E-05	3.3457E-05
3	4.9168E-05	4.95907E-05
4	6.52444E-05	6.57639E-05
5	8.1358E-05	8.1949E-05
6	9.75019E-05	9.81394E-05
7	0.000113674	0.000114333
8	0.000129872	0.000130527
9	0.000146097	0.000146723
10	0.000162348	0.00016292
11	0.000178624	0.000179117
12	0.000194926	0.000195315
13	0.000211253	0.000211513
14	0.000227606	0.000227711
15	0.000243983	0.000243909
16	0.000260386	0.000260107
17	0.000276814	0.000276306
18	0.000293268	0.000292504
19	0.000309746	0.000308703
20	0.00032625	0.000324902

and a warmup period increases as the correlation between the terms in the sum decreases, that is, if the correlation coefficient ρ approaches 1, the crossover point typically approaches infinity.

5. CONCLUSION

In this paper, we showed that if the MSE is to be minimized, and if a good starting state is known, then no initial transient deletion is required, and, in fact, any such deletion increases the MSE. Hence, in simulation, the effort should be geared toward finding a good starting state. There is always a starting state i^* that minimizes the MSE for a large enough simulation time T, but in our examples, there were usually a number of good starting states in the sense defined earlier. The problem thus becomes to find good starting states, which, I believe, will require a shift of the paradigm in the initialization bias theory. This paper is, of course, only a start in this direction, and we hope that others will follow. We believe that there are many results still waiting to discovered.

If warmup periods are used, one has to distinguish between starting state, the state the system starts with at time $-t_0$, and the time-zero state, the state when data collection begins. The objective of the warmup period is to reach, with high probability, a good time-zero state. This problem can often be solved by looking at the structural properties of the model, together with the entity to be estimated. For instance, if one wants to use simulation to find the output of the last queue in a sequence of queues, such as the output of an assembly line, then one obviously will not have a good time-zero state before the first job has a reasonable chance to reach the last server. In this connection, we note that in the M/D/∞ queue with a service time of 1, one can show that the bias disappears completely after simulating for 1 time unit (see [5] and references therein). This allows to

dispense with any statistical procedure otherwise needed to determine the length of the warm up period.

The bias can be substantial if one wants to estimate the expectation of a sum as indicated by considering equation (4). This is not surprising because in the time average of sums, the coefficient of variation decreases, that is, the stochastic component of the process in question decreases, which implies that it behaves more like a deterministic process. More generally, warmup periods become important when the sample function in question becomes more deterministic.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] H. P. Awad. On an initial transient deletion rule with rigorous theoretical support. In L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, editors, *Proceedings of the 2006 Winter Simulation Conference*, pages 186–191. The Institute of Electrical and Electronics Engineers, Inc., 2006.

[2] J. Banks, J. S. Carson, B. L. Nelson, and D. Nicol. *Discrete Event Simulation*. Prentice Hall, Englewood Cliffs, NJ, 2005.

[3] P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation*. Springer Verlag, New York, 1963.

[4] P. W. Glynn and D. L. Iglehart. A new bias detection rule. In A. Thesen, H. Grant, and D. Kelton, editors, *Proceedings of the 1987 Winter Simulation Conference*, pages 186–191. The Institute of Electrical and Electronics Engineers, Inc., 1987.

[5] W. K. Grassmann. Optimal estimation of the expected number in an M/D/∞ system. *Operations Research*, 29(6):1208–1211, 1981.

[6] W. K. Grassmann. Means and variances of time averages in Markovian environments. *European Journal of Operations Research*, 31:132–139, 1987.

[7] W. K. Grassmann. Warm-up periods in simulation can be detrimental. *Probabilitiy in Engineering and Informational Sciences*, 22:415–429, 2008.

[8] W. K. Grassmann and J. Luo. Simulating Markov-reward processes with rare events. *ACM Transactions on Modelling and Computer Simulation*, 15(2):138–154, 2005.

[9] J. Joo and J. L. Lebowitz. Pair approximation of the stochastic susceptible-infected-recovered-susceptible epidemic model on the hypercube lattice. *Phys. Rev. E*, 70(3):36114–36123, 2004.

[10] A. M. Law and W. D. Kelton. *Simulation Modelling and Analysis*. McGraw Hill, New York, third edition, 2000.

[11] A. Madansky. Optimal conditions for a simulation problem. *Operations Research*, 24:572–577, 1976.

[12] D. McNickle, G. Ewing, and K. Pawlikowski. Transient deletion and the quality of sequential steady-state simulation. In *Proceedings of the 21st European Conference on Modelling and Simulation*, Prague, Czech Republic, June 2007.

[13] K. Pawlikowski. Steady-state simulation of queueing processes: a survey of problems and solutions. *ACM Computing Surveys*, 22(2):123–170, 1990.

[14] K. Pawlikowski. Do not trust all simulation studies of telecommunication networks. In *International Conference on Information Networking (ICOIN 2003)*, pages 889–908, Jeju Island, Korea, February 2003.

[15] J. F. Reynolds. Some theorems on the covariance structure of Markov chains. *J. Appl. Prob.*, 9:214–218, 1972.

[16] W. H. Sanders et al. Möbius, model-based environment for validation of system reliability, availability, security and performance. `http://www.mobius.uiuc.edu/papers.html`, accessed June 11, 2007.

[17] W. Stewart. *Introduction to Numerical Solutions of Markov Chains*. Princeton University Press, 1994.

[18] W. Whitt. The efficiency of one long run versus independent replications in steady-state simulation. *Management Science*, 37(6):645–666, 1991.

[19] J. R. Wilson and A. A. B. Prisker. Evaluation of startup policies in simulation experiments. *Simulation*, 31:79–88, 1978.

[20] J. R. Wilson and A. A. B. Prisker. A survey of research on the simulation startup problem. *Simulation*, 31:55–58, 1978.