EFFICIENT TRUNCATION POINT ESTIMATION FOR ARBITRARY PERFORMANCE MEASURES

Mirko Eickhoff,
Donald C. McNickle,
Krzysztof Pawlikowski

University of Canterbury
Christchurch, New Zealand
E-Mail: m.eickhoff@cosc.canterbury.ac.nz

KEYWORDS

Sequential Steady State Simulation, Output Analysis, Truncation Point Detection, Arbitrary Performance Measures

ABSTRACT

In steady state simulation the output data collected during the transient phase often causes a bias in estimators of steady state performance measures. Common advice is to discard data collected during this transient phase. Finding an appropriate truncation point is a well-known and still not completely solved problem. In this paper we propose an improved algorithm for the determination of the truncation point for the output data sequence when its probability distribution reaches (approximately) steady state. The required run time of this algorithm is substantially improved without reducing the accuracy of the results. Because this algorithm is based on comparisons of empirical cumulative distribution functions, the truncation point is valid for an arbitrary steady state performance measure, such as the mean, variance or quantiles.

INTRODUCTION

In discrete-event simulation the sequence of output data represents a realization of a stochastic process $\{X_j\}_{j=1}^{\infty}$. The consecutive observations of this process are usually correlated and influenced by the initial state I of the system chosen by the analyst. Let $F_j(x|I) := \Pr[X_j \leq x|I]$ denote the cumulative distribution function of X_j. Assuming an ergodic system, $F_j(x|I)$ is converging towards $F(x) = \lim_{j\to\infty} F_j(x|I)$ which is called the marginal cumulative distribution function of the process $\{X_j\}_{j=1}^{\infty}$ in steady state. The primary concern of steady state simulation is to determine this distribution or its specific measures, such as e.g. the mean value.

In general the influence of I is significant in the beginning and decreases with increasing model time. If the interest is focused on the steady state behavior of the system, this initialization bias is obviously undesirable. A common way to reduce the influence of I is to truncate the "most" influenced part of the stochastic output process $X_1, \ldots, X_{l-1}$. Following this strategy the problem is to find an appropriate truncation point l. In the literature the steady state phase $\{X_j\}_{j=l}^{\infty}$ is described as a phase which is "relatively free of the influence of initial conditions" [Fishman, 2001] or by the statement that $X_l, X_{l+1}, \ldots$ "will have approximately the same (marginal) distribution" [Law and Kelton, 2000]. In practise there will often be an observation index l, such that

$$\forall j \geq l : F_j(x|I) \approx F(x) \tag{1}$$

is valid, unless the process $\{X_j\}_{j=1}^{\infty}$ is statistically unstable. Of course l should be finite, and should be the minimum of all indices, for which Equation (1) holds. Even though the estimation of $F(x)$ is the ultimate goal of steady state simulation, the expected value of the steady state random variable $E[X] = \lim_{j\to\infty} E[X_j]$ is often the only measure of interest. In this situation it is a generally accepted approach to replace Equation (1) by:

$$\forall j \geq l : E[X_j] \approx E[X] \tag{2}$$

In [Lee et al., 2000] it is shown that in the case of a $M/M/1/\infty$ system, the use of Equation (2) works even in quantile estimation. In general, however, the convergence of the mean is only a necessary condition for stationarity, and not a sufficient preliminary (see [Welch, 1983]). Therefore, Equation (1) can be applied in analysis of mean, variance, quantiles, or even rare event estimation, and Equation (2) should be used in mean value analysis only. In [Ghorbani, 2004] experimental investigations of methods based on Equation (2) are reported. Finding a truncation point on the basis of Equation (1) is not straightforward. It is therefore not very surprising, that the most common methods for detection of the truncation point are based on Equation (2) (see [Pawlikowski, 1990]) or on a visual inspection of the output data (see e.g. [Welch, 1983]). Completely algorithmic methods only give proper results under special conditions (see [Gafarian et al., 1978]) and are mostly based on one long simulation run. Nearly all methods use Equation (2), although the steady state phase should in general be defined by the convergence of $F_j(x|I)$ towards the steady state distribution $F(x)$.

In the MRIP scenario (multiple replications in parallel, see [Pawlikowski et al., 1994]) it is possible to collect a random sample of p independent and identical distributed realizations of each X_j, one from each replication. Let x_{ij} be the jth observation in the ith replication, with $1 \leq i \leq p$ and $1 \leq j < \infty$. Therefore, the empirical (marginal) cumulative distribution function $\tilde{F}_j(x|I)$ based on the order statistic

of $x_{1j}, \ldots, x_{pj}$ is an estimate of $F_j(x|I)$. This strategy requires a synchronization of the parallel replications so that the same number of observations is taken from each replication. Note, that this can lead to idle processors in an inhomogeneous computer network. The use of independent replications was considered in the case of mean value estimation in [Whitt, 1991] and [Alexopoulos and Goldsman, 2004]. The main source of error using independent replications is the initialization bias. If this source of error is eliminated, the estimates of independent replications are more accurate than e.g. ones obtained by means of methods based on batching. In [Eickhoff, 2002] an approach to reduce the initialization bias is proposed which is based on Equation (1) and uses the MRIP scenario. Its performance is examined and improved in [Bause and Eickhoff, 2002] and [Bause and Eickhoff, 2003]. An application of this approach is demonstrated in [Arns et al., 2003]. The results show that this approach is more reliable than methods which are based on Equation (2), especially if the transient behaviour is more complicated. Unfortunately, its computing time is quite large, possibly too excessive, as in many applications the computing time is as important as accuracy.

In the next sections an algorithm is described, which is based on Equation (1). In comparison to the algorithm described in [Bause and Eickhoff, 2003], the required run time of the new algorithm is substantially improved without reducing the accuracy of the results. This is examined in a later section by comparing the worst-case run time and the accuracy of both algorithms. In this paper only sequential methods and algorithms which are based on a dynamic set of data are considered. Therefore, the output analysis is performed online and guides the simulation experiment until estimates are statistically accurate.

IMPROVED ALGORITHM

The basic idea of the algorithm is given by Equation (1). The aim is to determine the first index l from which on all following probability distribution functions are (approximately) identical. Obviously the time horizon of every simulation experiment is limited. Therefore it is not possible to access "all" successive probability distribution functions. Accessible is only the observed part of the steady state phase. It is strongly necessary that this observed part of the steady state phase is reasonably large to avoid the determination of a misleading truncation point. Therefore the algorithm selects the size of the observed part of the steady state as a (constant) factor r of the size of the transient period, i.e. the size of the observed part of the steady state phase is always r-times larger than the so far selected transient phase. During each step of the algorithm the index l is increased by one. To assure the proper ratio between the two phases during each step $r+1$ new observation indices have to be accessed and included in analysis. For details on the basic idea of this approach see [Bause and Eickhoff, 2002]

Listing 1 shows a pseudo code of the improved algorithm, where for convenience some special notation is used. Let the sequence $\{y_{ij}\}_{i=1}^{p}$ be the order statistic of observations $\{x_{ij}\}_{i=1}^{p}$. Using the operators +, -, / and := in conjunction with sequences (see lines 0, 5, 8, 10 and 15) means to use these operators on each component of a given sequence separately. The operator $\approx$ in line 10 and 15 implements Equation (1) and is realized by the Kolmogorov-Smirnov two-sample test (KS-test). If the null hypothesis of equality is accepted, the operator $\approx$ (resp. the KS-test) results *true*. The procedure *observe()* collects one observation of each replication and the procedure *uniform(a,b)* delivers a uniform distributed integer random number between *a* and *b* used as index. The variable l points at X_l as the candidate for the truncation point, at the end of the algorithm it is the valid truncation point. The variable n represents the number of observations collected of each replication so far.

Listing 1: Pseudo code of the improved algorithm

```
0  int l := 0; int n := 0; int r := 10; {s_i}_{i=1}^p := 0;
   bool NoTestFailed := false;
   while (¬NoTestFailed){
      n := n + 1;
      observe({x_in}_{i=1}^p);
5     {s_i}_{i=1}^p := {s_i}_{i=1}^p + {y_in}_{i=1}^p;
      if (0 ≠ n mod (r+1)) continue;
      l := l + 1;
      {s_i}_{i=1}^p := {s_i}_{i=1}^p - {y_il}_{i=1}^p;
      NoTestFailed := true;
10    if (¬({y_il}_{i=1}^p ≈ {s_i}_{i=1}^p/(n-l)))
         NoTestFailed := false;
      for (int k := 1; k ≤ r; k := k+1){
         if (¬NoTestFailed) break;
         int u := uniform(lk+1, l(k+1));
15       if (¬({y_il}_{i=1}^p ≈ {y_iu}_{i=1}^p))
            NoTestFailed := false;
      }
   }
```

The most time consuming factor of the algorithm described in [Bause and Eickhoff, 2003] is the increasing number of KS-tests executed during each step of the algorithm. The improved algorithm described in Listing 1 avoids these tests by using the calculated sequence $\{s_i\}_{i=1}^{p}$ which is an estimate of $F(x)$ based on the latest observations during each step of the algorithm. The sequence $\{s_i\}_{i=1}^{p}$ is the sum of all order statistics which are not part of the transient period. New observations are added whereas observations of the transient period are subtracted from $\{s_i\}_{i=1}^{p}$ (see lines 5 and 8). Dividing each component of $\{s_i\}_{i=1}^{p}$ by the number of addends results in an estimate of $F(x)$. This sequence is compared with the order statistic of the actual test sample $\{x_{il}\}_{i=1}^{p}$ (see line 10). Because $\{s_i\}_{i=1}^{p}$ is calculated over observations at different model times, a possible periodic behaviour could be overlooked and an unreliable estimate of the truncation point could be accepted (cf. [Bause and Eickhoff, 2003]). Therefore additional r randomly chosen sequences are used to avoid this trap. The observed part of the steady state phase is divided into r equally sized intervals. Each interval contains one randomly chosen sequence. In a loop all of these sequences are compared with the actual test sample $\{x_{il}\}_{i=1}^{p}$ (see lines 12 to 17). If the assumption of equality is rejected by the KS-test for $\{s_i\}_{i=1}^{p}$, or any of the randomly chosen test samples, the truncation point l is not adequate and more steps of the algorithm have to be performed. In contrary to the previous version of this algorithm, the number of needed KS-tests in each step is maximum $r+1$.

WORST CASE TIME COMPLEXITY

In [Bause and Eickhoff, 2003] it is demonstrated experimentally, having considered a number of different kinds of transient

behaviour, that the previous version of the algorithm described in Listing 1 is very accurate and has therefore a great advantage over some other commonly used methods for truncation point estimation. But the price for its accuracy is the length of its run time which is $O(n^2 p \log(p))$. This run time is possibly too large for practical implementations. In this section it will be shown that the algorithm described in the previous section is a substantial improvement, because its run time is only $O(np\log(p))$, without significant loss of accuracy, in conjunction with any kind of transient behaviour.

As before, let p denote the number of replications and n the number of observations in each single replication. Note, that at the end of the algorithm n is a multiple of $r+1$. The total number of observations is pn. Assume, that all basic arithmetic operations are in $O(1)$ (cf. [Cormen et al., 1994]). In the following the running times of algorithms are considered from the point of view of their worst-case time complexity.

Theorem
The worst case running time of the algorithm described in Listing 1 is $O(np\log(p))$.

Proof: Only KS-tests with random samples of size p are performed. The basis of the KS-test are two sorted random samples. Sorting can be done in $O(p\log(p))$. To determine the maximum distance in the compared samples a pointer has to be shifted through each sorted random sample. This can be done in $2p$ steps which leads to $O(p)$. To accept or reject the null hypothesis, the determined maximum difference has to be compared with a tabulated critical value. This can be done in $O(1)$. Therefore, the run time of one KS-test is $O(p\log(p)) + O(p) + O(1) = O(p\log(p))$.

The run time of a single execution of lines 3 and 6 is $O(1)$ and of lines 4 and 5 it is $O(p)$. Because the while-loop in line 2 is executed n times before the algorithm stops, the run time of this part of the algorithm is $O(np)$.

A single execution of lines 7, 9 and 11 can be done in $O(1)$; a single execution of line 8 can be done in $O(p)$; to execute line 10 a run time of $O(p\log(p))$ is needed, because a KS-test has to be performed. Because of the condition in line 6 this part of the algorithm is executed only $\frac{n}{r+1}$ times. Therefore, the run time of this part of the algorithm is $\frac{n}{r+1} \cdot O(p\log(p))$ which leads to $O(np\log(p))$ because r is a constant parameter.

A single execution of lines 12, 13, 14 and 16 needs only a minor run time of $O(1)$. The KS-test in line 15 can be done in $O(p\log(p))$. The for-loop is executed at maximum r times, therefore, the run time of one complete for-loop in each step of the algorithm is $r \cdot O(p\log(p))$. All in all $\frac{n}{r+1}$ for-loops have to be performed. Therefore, the run time of this part of the algorithm is $n \cdot \frac{r}{r+1} \cdot O(p\log(p))$ which leads to $O(np\log(p))$.

Combining all results, the run time of the algorithm is $O(np) + O(np\log(p)) + O(np\log(p)) = O(np\log(p))$ ■

Because p could be considered as a constant parameter and usually $p << n$ holds, the run time could be described by $O(n)$. This run time is linear and highly efficient, because each observation has to be processed at least once.

PERFORMANCE

To compare the accuracy of both versions of the algorithm the same artificial processes as in [Bause and Eickhoff, 2003] are used, except the starting condition of the ARMA process is changed, to create a more unusual initial state. The output data of these artificial processes is used to show the performance of the algorithm on many different kinds of transient behaviour. Let $\{\epsilon_t\}_{t=1}^{\infty}$ denote an independent Gaussian white noise process (see [Hamilton, 1994]). In all experiments we used $p = 100$, $r = 10$ and an α-level of 0.05 for the critical value of the KS-test. For details on this parameters the reader is referred to [Bause and Eickhoff, 2002]. Further more, in all of the performed experiments the pseudo random number generator described in [L'Ecuyer et al., 2002] is used. This generator is suitable for many parallel replications, because its period is reasonably large and can be divided into many substreams.

Process A: linear transient mean

$$Y_t^{(A)} = \begin{cases} \epsilon_t + x - t\frac{x}{l} & \text{if } t < l, \\ \epsilon_t & \text{else.} \end{cases}$$

with $x = 10$, $l = 100$. This process can be regarded as the easiest case, because all performance measures are influenced by the initial state and show a transient behaviour, especially the mean and all quantiles. At a well defined index l the influence of the initial condition disappears completely.

Process B: linear transient variance

$$Y_t^{(B)} = \begin{cases} \epsilon_t \cdot (x - t\frac{x-1}{l}) & \text{if } t < l, \\ \epsilon_t & \text{else} \end{cases}$$

with $x = 10$, $l = 100$. The probability distribution function of this process has a transient behaviour, which is not visible for Equation (2) because the mean (resp. the median) is constant right from the beginning. Therefore, in this case Equation (2) is not suitable for an estimation of any other measure than the mean.

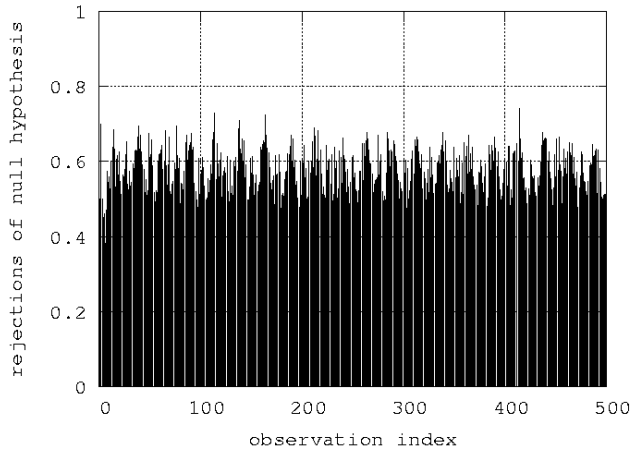
Process C: exponential transient mean

$$Y_t^{(C)} = \epsilon_t + x \cdot e^{(t\frac{\ln(0.05)}{l})}$$

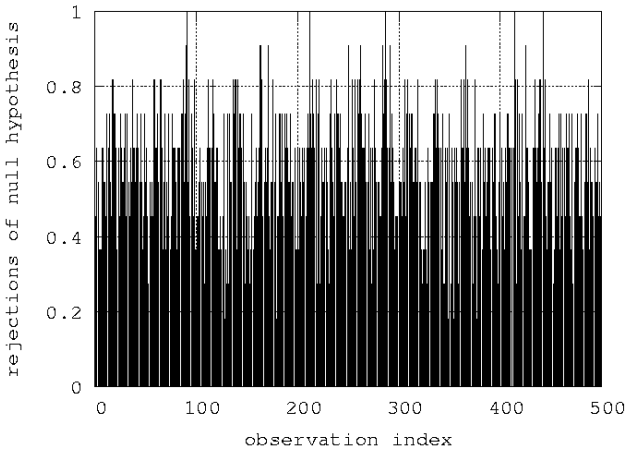
with $x = 10$, $l = 100$. In this case every performance measure has a transient behaviour. However, the influence of the initial state is disappearing exponentially. When regarding a finite simulation horizon, the influence never disappears completely. The definition of this process implies that at index l the impact of the influence of the initial state is reduced to 5% of its impact at $t = 0$.

Process D: ARMA$(5, 5)$

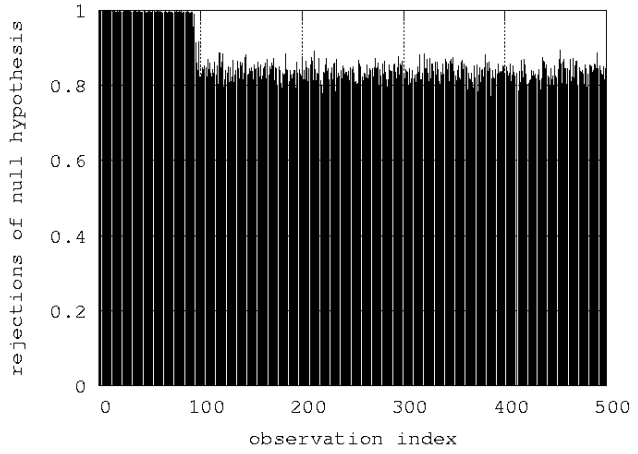
$$Y_t^{(D)} = 1 + \epsilon_t + \sum_{i=1}^{5} \frac{1}{2^i}(Y_{t-i}^{(D)} + \epsilon_{t-i})$$



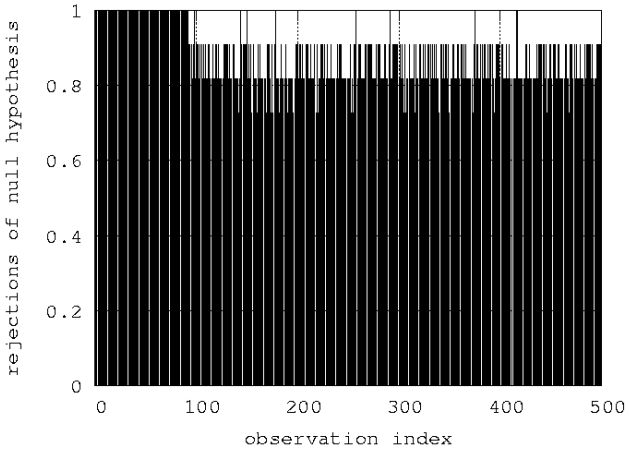
(a) Process E / previous algorithm



(b) Process E / improved algorithm



(c) Process F / previous algorithm



(d) Process F / improved algorithm

Figure 1: The number of rejections of the null hypothesis standardized by the number of all performed KS-tests. The values are plotted over model time, each peak represents one step of the algorithms.

with $Y_{-5}^{(D)} = Y_{-4}^{(D)} = Y_{-3}^{(D)} = Y_{-2}^{(D)} = Y_{-1}^{(D)} = 100$. Each value of this process depends on the 5 previous values. Therefore, the observations of this process are highly autocorrelated. A theoretical investigation shows that the expected value of this process in steady state is $E[Y_{\infty}^{(D)}] = 32$. The initial state has no influence on this expected value.

Process E: periodic

$$Y_t^{(E)} = \epsilon_t + b \cdot \sin(\omega t)$$

with $b = 1$, $T = \frac{2\pi}{\omega} = 50$. The transient behaviour of this process is governed by a sine oscillation and is not converging to a steady state distribution at all. A possible pitfall concerning output analysis of this process is that a batched mean seems to converge, if the batch size is equal to the period of the oscillation.

Process F: non-ergodic

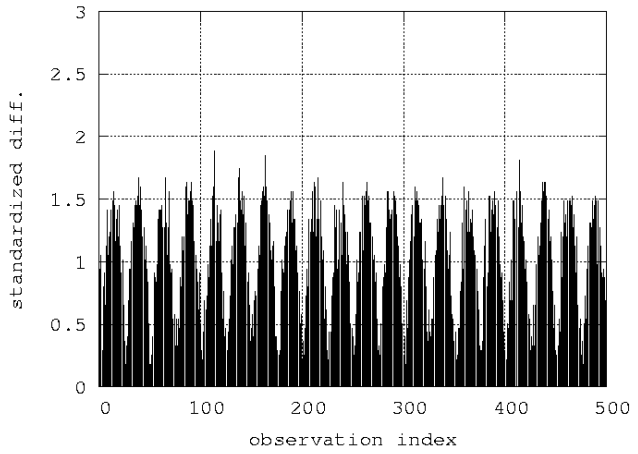
$$Y_t^{(F)} = \begin{cases} \epsilon_t(ct+1) + x - t\frac{x}{l} & \text{if } t < l, \\ \epsilon_t(ct+1) & \text{else.} \end{cases}$$

with $x = 10$, $l = 100$, $c = 0.01$. This process is governed by two different transient behaviours. The first one is obvious and is the same as in Process A. The second transient behaviour is more hidden and results in a non-ergodic behaviour. The pitfall concerning output analysis of this process is to overlook the second transient behaviour. This is especially likely when performing a visual inspection.

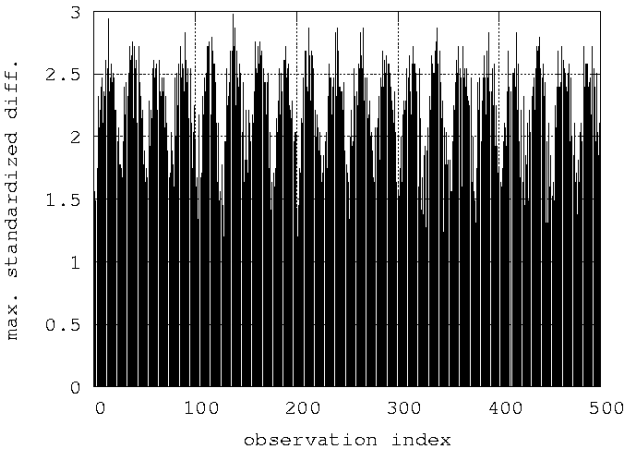
The processes A to D converge towards a steady state distribution. Therefore Equation (1) can be used to estimate the beginning of the steady state phase. To obtain a statistically accurate result simulation experiments with the algorithm described in [Bause and Eickhoff, 2003] and the algorithm of Listing 1 are performed 20 times. The average of all 20 results is listed in Table 1 with the halfwidth of its confidence interval and its relative statistical error.

using 20 runs	previous algorithm	improved algorithm
Process A	98.00 ± 0.43 (0.4%)	97.25 ± 0.57 (0.6%)
Process B	86.20 ± 1.01 (1.2%)	82.20 ± 1.87 (2.3%)
Process C	103.05 ± 1.76 (1.7%)	99.25 ± 1.81 (1.8%)
Process D	190.25 ± 6.53 (3.4%)	185.00 ± 6.55 (3.5%)
run time	$O(n^2 p \log(p))$	$O(np \log(p))$

Table 1: The average truncation points and their confidence intervals. Each result is based on 20 simulation experiments.



(a) Process E / Line 10: KS-tests with $\{s_i\}_{i=1}^{p}$



(b) Process E / Line 15: KS-tests with r random samples

Figure 2: Standardized statistics produced by the KS-tests executed in line 10 an line 15 respectively (see Listing 1) when applied to Process E. The null hypothesis is accepted when the standardized statistic is below 1.

The halfwidths of the confidence intervals are in all our experiments smaller than five percent of the mean value after fewer than 20 simulation experiments. This shows that both algorithms deliver a robust estimate with a small variance. Even though the results of both algorithms are comparable, the algorithm described in Listing 1 tends to estimate the truncation point a bit earlier. This shows that the fewer KS-tests of this algorithm are weaker than the large number of KS-tests in the previous version of the algorithm, which cause the longer run time.

The processes E and F do not converge towards a steady state distribution, and so there is no steady state phase at all. Therefore, both algorithms should not return a truncation point. To check the accuracy in these cases each step of the algorithms is observed by plotting the number of rejections of the null hypothesis (see Figure 1). The rejections should be on a high level so that Equation (1) will not be true for any tested value of l. To make the number of rejections comparable, they are standardized by the number of all performed KS-tests. Therefore, the value 1 means that all KS-tests reject the null hypothesis, and the value 0 means that all KS-tests accept the null hypothesis. To obtain results depicted in Figures 1(b), 1(d) and 2(b) the condition in line 13 of Listing 1 is ignored, to force the execution of all $r + 1$ KS-tests to obtain continuous plots.

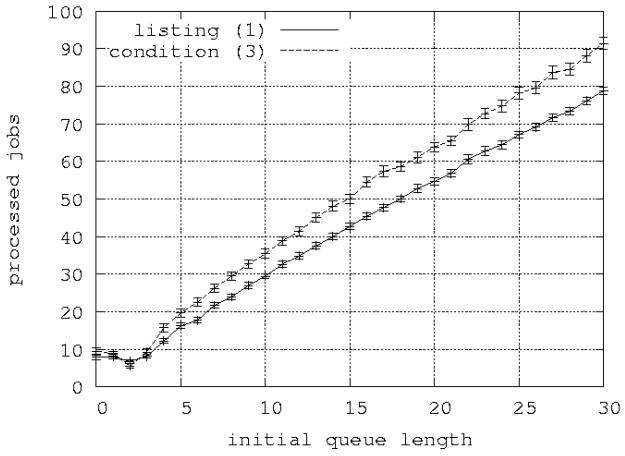
The plots in Figure 1(a) and 1(c), with the results for the original algorithm described in [Bause and Eickhoff, 2003] are quite smooth, because a large number of KS-tests are used to check Equation (1). Additionally these plots show that the KS-tests of the previous algorithm work reliably, because of the large number of rejected null hypotheses when analyzing Process E and F. Note, this algorithm gives an estimated length of the initial transient phase if the number of rejections is below a certain threshold, which is usually set at 0.05. The plots in Figure 1(b) and 1(d), for the modified algorithm, are not as smooth, because the number of executed KS-tests is much smaller. However, the number of rejections of the null hypothesis is high in both cases. Note, that the improved algorithm accepts an estimate only if all KS-test accept the null hypothesis. Therefore the improved algorithm works in the case of on Process E and F as reliably as the previous algorithm.

Furthermore, Process E is a good example to demonstrate that a truncation point estimation exclusively based on $\{s_i\}_{i=1}^{p}$ is not sufficient. The statistics produced by the KS-test are plotted in Figure 2. Figure 2(a) shows the results of the comparisons between $\{x_{il}\}_{i=1}^{p}$ and $\{s_i\}_{i=1}^{p}$ performed in line 10 of Listing 1. Figure 2(b) shows the maximum statistics of all KS-tests performed in line 15 of Listing 1 during the for-loop. In the second case, the maximum statistics of all r comparisons are plotted by disregarding the condition from line 13. To achieve comparability, the statistics are standardized by appropriate critical values of the KS-test. A standardized statistic below 1 means that the corresponding null hypothesis was accepted by the KS-test.

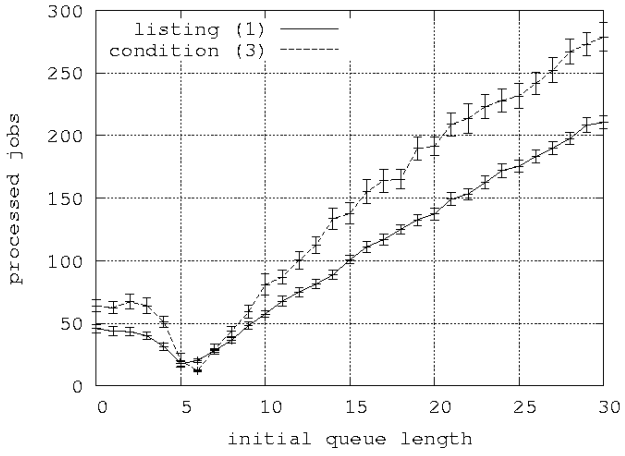
Figure 2(a) shows a periodical pattern of acceptance of the null hypothesis. The reason for this is, that $\{x_{il}\}_{i=1}^{p}$ is governed by the periodical behaviour of Process E. The situation is different in the case of $\{s_i\}_{i=1}^{p}$ since then many random samples of different model times are taken into account. Therefore, $\{s_i\}_{i=1}^{p}$ cannot reflect the periodical behaviour and represents the average test sample over many periods of the sine oscillation. When Process E is at an extreme value of the sine oscillation, the KS-test detects the difference between $\{x_{il}\}_{i=1}^{p}$ and $\{s_i\}_{i=1}^{p}$. If Process E is at a zero point of the sine oscillation, the KS-test is not able to recognize a difference between $\{x_{il}\}_{i=1}^{p}$ and $\{s_i\}_{i=1}^{p}$. These results in the same problem as shown for the method of Welch (see [Bause and Eickhoff, 2003] and [Welch, 1983]). However, the KS-tests with the randomly selected samples make sure that the improved algorithm works reliably (see Figure 2(b)).

TRANSIENT BEHAVIOUR OF THE M/M/1 QUEUE

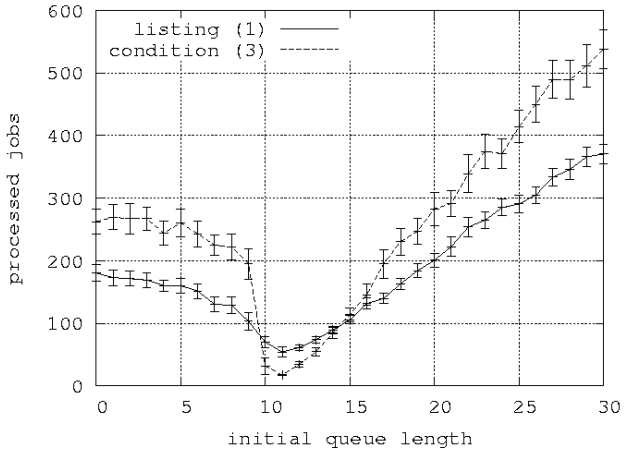
In [Kelton and Law, 1985] it is pointed out that true steady state behaviour of a M/M/1 queueing system is very difficult to observe by simulation experiments. The reason is that the dependence of the output data on the initial state can last very long. In [Jain, 1991] the most important properties of an M/M/1 queue are summarized. Only the mean arrival rate λ and the service rate μ is necessary to analyze the steady state behaviour of the



(a) traffic intensity $\rho = 0.5$



(b) traffic intensity $\rho = 0.8$



(c) traffic intensity $\rho = 0.9$

Figure 3: The estimated truncation points in response time analysis obtained according to Listing 1 and Condition (3).

M/M/1 queue. The traffic intensity is given by $\rho = \frac{\lambda}{\mu}$. The stability condition is $\rho < 1$. Let N denote the number of jobs in the system in steady state, in queue and in service. If the stability condition holds, the mean number of jobs in the system in steady state is $E[N] = \frac{\rho}{1-\rho}$. Let D denote the response time (or turnaround time) of the system in steady state, including the time in queue and in service. Then, the mean response time in steady state is given by $E[D] = E[N]/\lambda$. Let D_{ij} denote the response time of the jth job leaving the system in the ith replication. When using the MRIP scenario, the mean response time of the jth job, $E[D_j]$, can be estimated by

$$\bar{D}_j(p) = \frac{1}{p}\sum_{i=1}^{p} D_{ij} \quad .$$

The aim of the experiments reported here is to assess the results produced by the algorithm described in Listing 1 when analyzing the M/M/1 queue. A comparison with the previously published results of [Kelton and Law, 1985] is done. The authors of that publication used the same condition as a measure of convergence to steady state, as [Gafarian et al., 1978]. Namely:

$$|\bar{D}_j(p) - E[D]| \leq \epsilon \cdot E[D] \tag{3}$$

When analyzing the response time D, j should not be smaller than the initial number of jobs in the system. However, it should be the minimum index for which this condition holds. This condition is based on the estimated mean values $\bar{D}_j(p)$ and, therefore, implements Equation (2). Further more, this condition is only applicable, if the expected steady state value of $E[D]$ is known. Therefore, this rule has only theoretical importance and cannot be applied in practise.

The primary aim in [Kelton and Law, 1985] was to use that condition to assess the rate of convergence of the mean waiting time in the queue in different queueing models to steady state. Because of this, the results of the algorithm described in Listing 1 are not expected to be the same as those of Kelton and Law.

In all our experiments we set $\lambda = 1$ so that $E[D] = E[N]$ holds. The series of experiments are done for estimating the truncation point of the M/M/1 queue with a constant traffic intensity of $\rho = \{0.5; 0.8; 0.9\}$ and an initial queue lengths between 0 and 30. In this experiments we set the parameter $r = 1000$. This leads to a large ratio between the length of the transient phase and the observed part of the steady state phase when executing the modified truncation point detection algorithm. For each initial value of the queue length the experiment is repeated 100 times to estimate the average truncation point and its confidence interval. In parallel to the algorithm described in Listing 1, Condition (3) is being checked on the same set of output data. We set $\epsilon = 0.05$ for all experiments. The analyzed measure is the response time.

The results are plotted in Figure 3 for $\rho = \{0.5; 0.8; 0.9\}$. The shapes of the curves obtained under Condition (3) are similar to those obtained for the improved truncation algorithm. In all cases they start on a higher level and drop down at the point, where the initial queue length is (approximately) equal to $E[N]$ resp. $E[D]$. After this point all curves rise again. This shows, that the results of the improved algorithm conform the theory.

The convergence of the mean is a necessary condition for a process being in the steady state phase. Therefore, a truncation point chosen on the basis of the converging probability distribution is expected to be equal or larger than a truncation point chosen on the basis of the converging mean. However, in the implementation of Equation (1) and Condition (3) two different measures of convergence have been used, so the results are not directly comparable. One can see that the estimated value of the truncation point depends not only on the convergence of the observed measure, but also the measure of closeness which in

Condition (3) is represented by a scalar value ϵ, while in the implementation of Equation (1) the statistic of the KS-test is used.

The experiments with this M/M/1 queue show that the results of the algorithm described in Listing 1 conform to theory and are comparable to the results of [Kelton and Law, 1985]. In contrast to the previously published results, all simulation experiments were performed automatically without any a priori knowledge about the steady state distribution or steady state mean values. Note, that the large number of experiments at high traffic intensities is only feasible because of the efficient running time of the improved algorithm.

LIMITS AND REJECTED IDEAS

Because Equation (1) uses an approximation the estimate of the beginning of the steady state phase of Process A is always smaller than the theoretically best value which is in this example $l = 100$. However, using the equality instead of an approximation does not make sense in output processes such as Process C. Using the equality instead of an approximation would lead to an infinite value for l, because in Process C the influence of the initial state disappears exponentially.

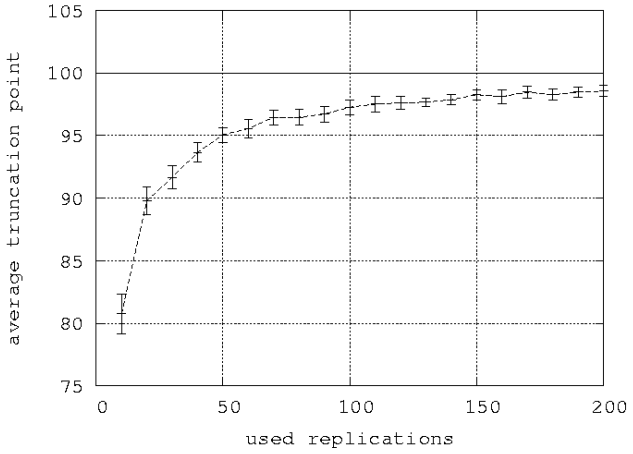


Figure 4: Truncation point analysis of Process A. The quality of estimates depends on the number of parallel replications.

Figure 4 depicts the results of simulation experiments for the Process A executed with different numbers of parallel replications. These experiments were executed as described in previous sections, with 20 independent simulation experiments for each plotted point. These results validate the assertion, that the estimated length of the initial transient phase is always smaller than its theoretically best value. Furthermore, another effect can be observed. The values of the estimates decrease with decreasing numbers of parallel replications. Note, that the critical values of the KS-test are defined for very small sample sizes, too. The realization of the approximation of Equation (1) based on the KS-test gets weaker for smaller numbers of parallel replications. Therefore, we recommend at least 30 parallel replications, and if possible more than 50 parallel replications should be used. This limit is valid for the algorithms described in [Bause and Eickhoff, 2003] and in Listing 1.

Before investigating the performance of the algorithm described in Listing 1, it was uncertain, whether the obtained results would be accurate or not. Therefore, it was an additional idea to perform a large number of KS-tests, as it is done in the previous version of the algorithm, whenever the $r+1$ KS-tests of the new version of the algorithm do not recognize a difference. This idea has been rejected because of two reasons. Firstly, the results obtained for both versions of the algorithm are convincingly close, see Table 1. Another reason for the rejection of that idea was, that the worst case time complexity would still be $O(n^2 p \log(p))$. In the worst case, the $r + 1$ KS-tests of the new version of the algorithm would always accept the null hypothesis and the large number of KS-tests of the previous version of the algorithm would always reject the null hypothesis. In this situation the worst case run time of the algorithm would be even worse than in the previous version.

CONCLUSIONS

We introduced an efficient algorithm for the estimation of the length of the initial transient phase in the sense of probability distribution; see Equation (1). The worst case time complexity of this algorithm is limited by $O(np \log(p))$ and is, therefore, substantially faster than the previous version of the algorithm (see [Bause and Eickhoff, 2003]). This improvement is achieved without reducing the accuracy of the estimates. Its performance has been experimentally assessed by conducting a number of simulations of artificial output processes with a variety of different types of transient behaviour.

We also investigated the newly introduced algorithm in M/M/1 queue simulation and compared them with previous results published in [Kelton and Law, 1985]. The results are as expected and conform to theory.

In the introduction it is pointed out that a truncation point estimation based on Equation (1) can be used for many different performance measures, such as mean values, variances, quantiles or even rare-events. Equation (2) is in general only useful in mean value analysis.

However, limits concerning Equation (1) are also discussed. For Process A is demonstrated that this equation does not always leads to the theoretically best truncation point. This is caused by the approximation used. Using the equality instead of an approximation is no alternative, because this would lead in some cases (e.g. Process C) to an infinite truncation point. Equation (1) can be used to reduce the initialization bias dramatically, but it cannot be used to eliminate it completely.

REFERENCES

[Alexopoulos and Goldsman, 2004] Alexopoulos, C. and Goldsman, D. (2004). To batch or not to batch. *ACM Transactions on Modeling and Computer Simulation*, 14(1):76–114.

[Arns et al., 2003] Arns, M., Eickhoff, M., Fischer, M., Tepper, C., and Völker, M. (2003). New Features in the ProC/B Toolset. In F. Bause (ed): Tools of the 2003 Illinois International Conference on Measurement, Modelling and Evalu-

ation of Computer-Communication Systems. Technical Report 781, Uni Dortmund, Fachbereich Informatik.

[Bause and Eickhoff, 2002] Bause, F. and Eickhoff, M. (2002). Initial transient period detection using parallel replications. *Proc. of the 14th European Simulation Symposium*, pages 85–92.

[Bause and Eickhoff, 2003] Bause, F. and Eickhoff, M. (2003). Truncation point estimation using multiple replications in parallel. *Proc. of the Winter Simulation Conference*, pages 414–421.

[Cormen et al., 1994] Cormen, T. H., Leiserson, C. E., and Rivest, R. L. (1994). *Introduction to Algorithms*. MIT Press.

[Eickhoff, 2002] Eickhoff, M. (2002). Statistische Auswertung und Erkennung der stationären Phase in der Simulation zustands-diskreter Systeme. Diploma Thesis, Uni Dortmund, Fachbereich Informatik.

[Fishman, 2001] Fishman, G. S. (2001). *Discrete-Event Simulation*. Springer.

[Gafarian et al., 1978] Gafarian, A. V., Ancker, C. J., and Morisaku, T. (1978). Evaluation of commonly used rules for detecting steady state in computer simulation. *Naval Research Logistics Quarterly*, pages 511–529.

[Ghorbani, 2004] Ghorbani, B. (2004). The issue of initial transient in sequential steady-state simulation. Master thesis, University of Canterbury, Department of Computer Science.

[Hamilton, 1994] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

[Jain, 1991] Jain, R. (1991). *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, Inc.

[Kelton and Law, 1985] Kelton, W. D. and Law, A. M. (1985). The transient behavior of the m/m/s queue, with implications for steady-state simulation. *Operations Research*, 33:378–396.

[Law and Kelton, 2000] Law, A. M. and Kelton, W. D. (2000). *Simulation Modeling and Analysis*. McGraw-Hill Higher Education, New York.

[L'Ecuyer et al., 2002] L'Ecuyer, P., Simard, R., Chen, E. J., and Kelton, W. D. (2002). An object-oriented random-number package with many long streams and substreams. *Operations Research*, 50(6):1073–1075.

[Lee et al., 2000] Lee, J.-S. R., McNickle, D., and Pawlikowski, K. (2000). Initial transient period detection for steady-state quantile estimation. *Proceedings of the Summer Computer Simulation Conference*, pages 169–174.

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

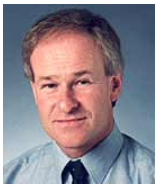
[Pawlikowski et al., 1994] Pawlikowski, K., Yau, V., and McNickle, D. (1994). Distributed stochastic discrete-event simulation in parallel time streams. *Proc. of the Winter Simulation Conference*, pages 723–730.

[Welch, 1983] Welch, P. D. (1983). The statistical analysis of simulation results. *In The Computer Performance Modeling Handbook, ed. S. Lavenberg, Academic Press*, pages 268–328.

[Whitt, 1991] Whitt, W. (1991). The efficiency of one long run versus independent replications in steady-state simulation. *Management Science*, 77(6):645–666.

AUTHOR BIOGRAPHIES

MIRKO EICKHOFF holds a Diploma degree in Computer Science from the University of Dortmund. His research interests are in the area of output analysis of discrete event simulation using multiple replications. His diploma thesis is part of the Collaborative Research Center "Modelling of Large Logistic Networks" (559) supported by the Deutsche Forschungsgemeinschaft. He worked for Delmia (Germany) in the area of workload balancing in manufacturing industry. In 2004 he received the targeted doctoral scholarship of the University of Canterbury and is currently a Ph.D. Candidate in Computer Science in the Simulation Research Group at this University. His e-mail address is m.eickhoff@cosc.canterbury.ac.nz.



DONALD C. MCNICKLE is an Associate Professor of Management Science in the Management Department at the University of Canterbury. His research interests include queueing theory, networks of queues and statistical aspects of stochastic simulation. He is a member of INFORMS and the Operational Research Society. His e-mail address is don.mcnickle@canterbury.ac.nz.

KRZYSZTOF PAWLIKOWSKI is a Professor in Computer Science at the University of Canterbury, in Christchurch, New Zealand. The author of over 130 research papers and four books; has given invited lectures at over 80 universities and research institutes in Asia, Australia, Europe and North America. Alexander-von-Humboldt Research Fellow (Germany) in 1983-84 and 1999. His research interests include performance modelling of telecommunication networks, discrete-event simulation and distributed processing. Senior Member of IEEE, member of ACM and SMSI. His e-mail address is krys.pawlikowski@canterbury.ac.nz.