DETECTING THE DURATION OF INITIAL TRANSIENT IN STEADY STATE SIMULATION OF ARBITRARY PERFORMANCE MEASURES

Mirko Eickhoff
Computer Science
University of Canterbury
Christchurch, New Zealand
m.eickhoff@cosc.canterbury.ac.nz

Don McNickle
Management
University of Canterbury
Christchurch, New Zealand
don.mcnickle@canterbury.ac.nz

Krzysztof Pawlikowski
Computer Science
University of Canterbury
Christchurch, New Zealand
krys.pawlikowski@canterbury.ac.nz

ABSTRACT

The issue of the initial transient phase in steady state simulation has been widely discussed in simulation literature. Many methods have been proposed for deciding the duration of this phase of simulation, to determine a valid truncation point of the transient portion of output data. However, practically all these methods can only be used in simulations aimed at estimation of mean values. In this paper, we show that analyses of performance measures which do not represent mean values require different solutions, as the rate of convergence to steady state is different for mean values than, for example, for quantiles. We describe and present additional results for a new method of determining the duration of initial transient phase which can be applied in analysis of steady state quantiles and probability distributions. The method appears robust and applicable in analysis of arbitrary performance measures.

Categories and Subject Descriptors

I.6.6 [**Simulation and Modeling**]: Simulation Output Analysis; G.3 [**Probability and Statistics**]: Distribution Functions; Stochastic Processes

General Terms

Performance, Algorithms, Experimentation

Keywords

Discrete event simulation, steady state, initial transient, truncation point

1. INTRODUCTION

In every simulation experiment the initial state of the model has to be set. Assuming that the main focus of simulation is the long run behaviour of a simulated system, a good initial state would be a state, which is typical of the long run behaviour. However, because the long run behaviour of the model is unknown, this kind of initialisation is not possible in general. To reduce bias in the final estimate a common approach is to split the sequence of collected output data into two parts. The first part contains data collected during the initial transient phase and the remaining sequence represents the steady state phase of the simulated system's behaviour. These two parts are separated by a data item of index l, which is referred to as the truncation point of the initial transient phase. In most methods of simulation output data analysis, the output data from the first part of the sequence are discarded before the rest of the collected data sequence are used in analysis of steady state behaviour of the system.

The rate of convergence towards long run behaviour depends on the system itself, and also on the performance measure considered. Most of the methods for determining the truncation point l, which have been proposed so far, are valid in mean value analysis only. Here, the aim is to determine the truncation point l, which is valid for estimation of any steady state parameter, regardless of whether it is a mean, variance or quantile.

In the next section we will define a new class of truncation points by relating them to the type of analysed performance measures they were introduced for. In Section 3 we will demonstrate why this is needed by discussing, as an example, the transient behaviour of quantiles of the response time of an M/M/1 queue for different initial states. Then, in Section 4, we will show that some well known methods for the detection of steady state are not applicable in this case. We suggest a method for the detection of steady state in the sense of the underlying probability distribution. This method can be used for analysis of arbitrary performance measures. Results of our approach are discussed in Section 5 by calculating the probability distribution of the estimated onset of steady state. Some conclusions are given in Section 6. The discussion and results which are presented in this paper are an extension of [5] and an excerpt of [6].

Valuetools '07, October 23-25, 2007, Nantes, France

2. CONCEPT OF STEADY STATE

Practically all methods, which were proposed for estimating the duration of initial transient in simulation output data analysis, are applicable in mean value analysis only. Comprehensive surveys of these methods can be found in, for example, [17] and [18]. The output stream of a simulation run is a stochastic process $\{X_i\}_{i\in\mathbb{N}}$. In many simulation studies the objective is to analyse the system's behaviour in the long run or in steady state, i.e. to analyse X_i as $i \to \infty$. During the steady state phase, i.e. for $i \geq l$, the output data are assumed to represent of the system's steady state behaviour because they are (approximately) not influenced by the initial state. Thus,

$$\forall (i \geq l_F, \Delta \geq 0, x) : F_{X_i}(x) \approx F_{X_{i+\Delta}}(x) . \qquad (1)$$

We use l_F instead of l to explicitly point out that the truncation point is determined by inspecting $F_{X_i}(x)$. However, it is known that different performance measures, in particular different moments, converge to steady state at different rate. An empirical proof of this fact is shown in Section 3 of this paper. Thus, we can define the steady state phase in terms of the mean by

$$\forall (i \geq l_E, \Delta \geq 0) : \mathrm{E}[X_i] \approx \mathrm{E}[X_{i+\Delta}] . \qquad (2)$$

Equation (2) can be used if the only target of the simulation is to estimate $\mathrm{E}[X_\infty]$. We define the steady state phase in terms of the variance by

$$\forall (i \geq l_V, \Delta \geq 0) : \mathrm{Var}[X_i] \approx \mathrm{Var}[X_{i+\Delta}] . \qquad (3)$$

Other definitions may be appropriate, e.g. for the 0.95-quantile of the distribution of X_i the truncation point $l_{0.95Q}$ could be defined analogously to (2) and (3).

Constant first and second moments are necessary conditions for (1). Thus, the steady state phase of the output process must imply that the mean and the variance are in their steady state phase, i.e. $l_E \leq l_F$ and $l_V \leq l_F$. However, whether $l_E \leq l_V$ or $l_E \geq l_V$ depends on the properties of the output process. Note that it is possible to find processes for which (2) and/or (3) hold, but not (1). In this situation $\mathrm{E}[X_\infty]$ can be estimated even though $F_{X_\infty}(x)$ does not exist. The counterpart of the steady state phase is the transient phase with $i < l$. During the transient phase (1) does not hold.

In analysis of stochastic processes stationarity is an important property and is discussed e.g. in [16], [14] and [22]. A stochastic process $\{X(t)\}_{t\in T}$, not necessarily representing simulation output data, is stationary (in the strict sense) if its statistics are not affected by a shift in the time origin. This means that two processes $\{X(t)\}_{t\in T}$ and $\{X(t+\Delta)\}_{(t+\Delta)\in T}$ have the same statistics for any Δ. The joint distribution of any set of samples of a stationary process does not depend on the placement of the time origin:

$$F_{X(t_1),\ldots,X(t_j)}(\vec{x}) = F_{X(t_1+\Delta),\ldots,X(t_j+\Delta)}(\vec{x}), \qquad (4)$$

for all time shifts Δ, all j and all choices of sample times t_1, $\ldots$, t_j. If (4) is true, not for any j, but only for $j \leq k$ the process $\{X(t)\}_{t\in T}$ is stationary of order k. Therefore, the simulation output process $\{X_i\}_{i\geq l_F}$ can be assumed to be stationary of first order, if (1) can be fulfilled.

3. TRANSIENT BEHAVIOURS OF M/M/1 SYSTEMS

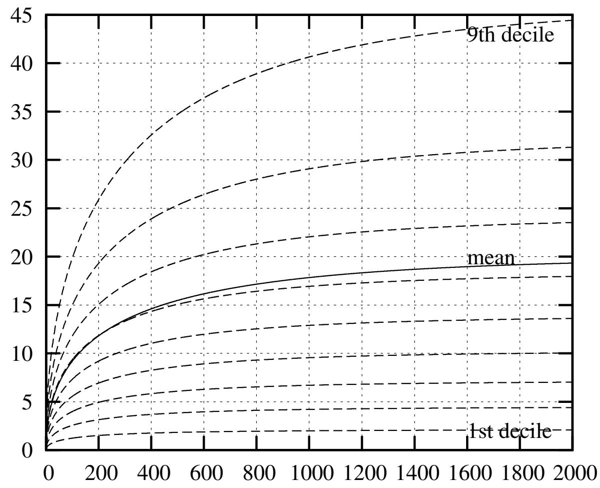
In this section we demonstrate the kinds of convergence of $F_{X_i}(x)$ towards $F_{X_\infty}(x)$ that may occur, by considering an M/M/1 queueing system. Its steady state behaviour and its transient behaviour are well known ([13], [15] and [12]). The distribution of the number of customers N_i in the system at the arrival time of the ith customer can be calculated by a Markov chain approach, imbedded at arrival times. Based on the distribution of N_i other measures can be calculated, e.g. the system's response time R_i (the sum of waiting time and service time) of the ith customer. The probability distributions, mean values and quantiles (or more specifically deciles) depicted in Figure 1 were calculated by the analytical approach.

First we study the transient behaviour of a M/M/1 server initialised with an empty queue and an idle server. This initialisation state is chosen very often, because it is this system's state with the highest probability for $\rho < 1$. We chose $\lambda = 0.95$ and $\mu = 1$ resulting in $\rho = \frac{\lambda}{\mu} = 0.95$. In Figure 1(a) the mean and the deciles of $F_{R_i}(x)$ are depicted, so that the ordinate shows the range of R_i and the abscissa shows i. The mean and the deciles are strictly monotonically increasing and converging. Even though the M/M/1 server is initialised with its state of highest probability a transient phase is present. We note in passing that the convergence of the mean value is similar to the convergence of the $6th$ decile. This is not surprising, as $F_{R_\infty}(\mathrm{E}[R_\infty]) = 1 - e^{-\mathrm{E}[R_\infty]\mu(1-\rho)} \approx 0.632$. Smaller deciles converge faster, higher deciles converge slower. However, the general form of convergence looks similar for the mean and all deciles. In Figure 1(b) the distribution $F_{R_i}(x)$ is plotted for selected values of i and for $i = \infty$, so that the ordinate shows the cumulative probability and the abscissa shows the range of R_i. The distribution $F_{R_i}(x)$ is a weighted sum of Erlang distributions and the limit response time distribution is exponential. Thus, in (rough) approximation we can regard $F_{R_i}(x)$ as an exponential distribution with increasing mean value as i is growing. Thus, the convergence can be loosely described as a scaling of the distribution.

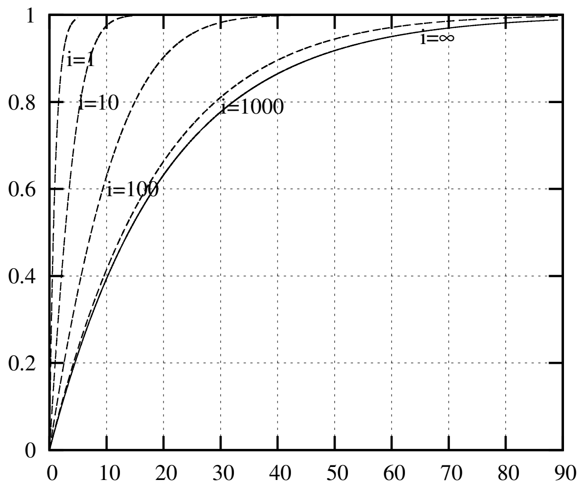
For our second example we use $\lambda = 0.95$ and $\mu = 1$. This gives $\mathrm{E}[N_\infty] = \frac{\rho}{1-\rho} = 19$ customers. Therefore, we chose 19 initial customers. This is depicted in Figures 1(c) and 1(d), including the response time of the initial customers ($i \leq 19$). Again, $F_{R_i}(x)$ converges towards $F_{R_\infty}(x)$. However, in Figure 1(c) we can see that in this case the convergence is not monotonic. This is because the form of $F_{R_{20}}(x)$, the response time of the first non-initial customer, is different from the form of $F_{R_\infty}(x)$. In Figure 1(d) we can see that the density of $F_{R_{20}}(x)$ is almost symmetric and $F_{R_\infty}(x)$ is an exponential distribution. Here, we cannot assume that the convergence can be described by a fixed class of distributions with an additional scaling or displacement. This is interesting especially for mean value analysis, as it cannot be assumed that the mean value is constant right from the beginning of a simulation if the system is initialised with $\mathrm{E}[N_\infty]$ customers. We note that our results conform to the observation of [13] that in some simple queues the optimal initial state for mean value analysis is higher than $\mathrm{E}[N_\infty]$.

In our last example we choose an initial state, which is much higher than $\mathrm{E}[N_\infty]$. We used $\lambda = 0.8$, $\mu = 1$ and one hundred initial customers of the M/M/1 server. In Figures 1(e) and 1(f) we can see that the density of $F_{R_{101}}(x)$, the response time of the first non-initial customer, is almost symmetric, i.e. clearly non-exponential. $F_{R_i}(x)$ converges until it is exponentially distributed at $i = \infty$. Again, no constant distribution can be assumed, and also there is an obvious displacement of the mean value during the transient phase.

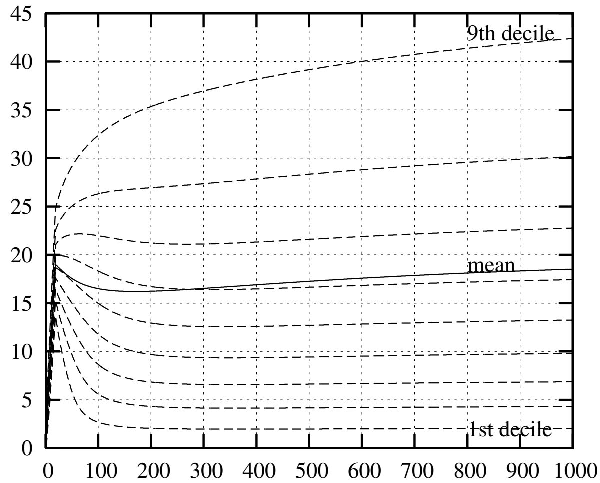
Thus, we see that the convergence of $F_{R_i}(x)$ towards $F_{R_\infty}(x)$ is complex, even in the case of the M/M/1 server.



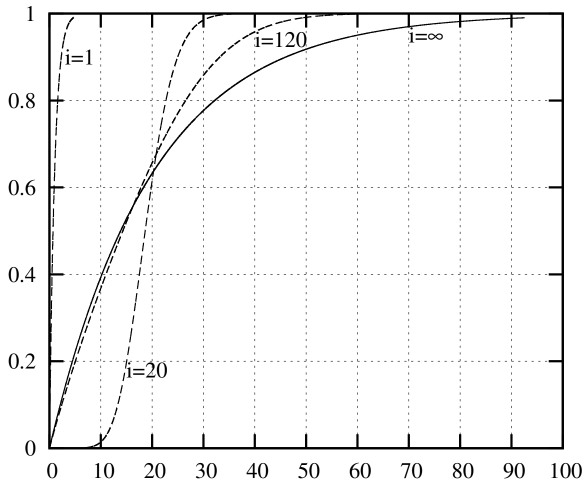
(a) Deciles: $\lambda = 0.95$, $\mu = 1$, empty and idle



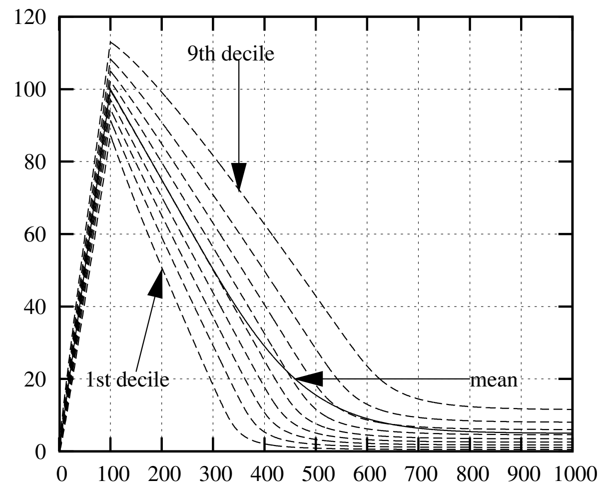
(b) $F_{R_i}(x)$: $\lambda = 0.95$, $\mu = 1$, empty and idle



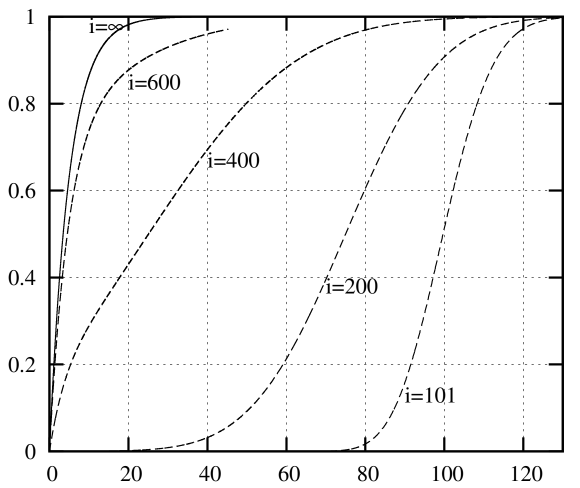
(c) Deciles: $\lambda = 0.95$, $\mu = 1$ and 19 initial customers



(d) $F_{R_i}(x)$: $\lambda = 0.95$, $\mu = 1$ and 19 initial customers



(e) Deciles: $\lambda = 0.8$, $\mu = 1$ and one hundred initial customers



(f) $F_{R_i}(x)$: $\lambda = 0.8$, $\mu = 1$ and one initial hundred customers

Figure 1: Theoretical convergence of deciles and cumulative distribution function of response times of the M/M/1 server with different initial states. Depicted is the evolution of deciles (dashed) and the mean (bold) and $F_{R_i}(x)$ of selected customers i (dashed) and for $i = \infty$ (bold).

For some i the density of $F_{R_i}(x)$ can be almost symmetric, whereas $F_{R_\infty}(x)$ is exponentially distributed. The form of the convergence depends heavily on the initialisation of the server. The convergence of the mean is not necessarily monotonic and can be either like or unlike the convergence of deciles.

4. CONSTRUCTION OF DETECTION METHODS

The main purpose of this section is to discuss why classical methods for the detection of a steady state mean are not applicable to detect steady state in terms of the underlying distribution function. To demonstrate that it is possible to implement a detection method that is based on (1) a brief discussion of this implementation is given.

Our interest focuses on automated and sequential analysis of arbitrary performance measures. Therefore, a detection method of the steady state in terms of (1) is needed because if steady state in the sense of distribution is reached then it is guaranteed that all other performance measures (moments, quantiles, etc.) have reached steady state as well. Thus, an estimator for l_F is needed. Detection of the steady state in terms of (2) or (3) is legitimate if the interest is on analysis of means or variances only. If the data is truncated at l_F the remaining data can be assumed to be approximately identically distributed. This is a very strong property and it is important for many statistical methods. In automated analysis we cannot require any previous knowledge of the output data. Thus, the detection method must be able to deal with a wide range of transient behaviours, including unstable models.

As we have demonstrated in Section 3, the transient behaviour of even a simple system can be quite complex. To construct an estimator of l_F that is valid for any output process a robust detection method is needed. We will discuss some detection methods that have been proposed for determining l_E and demonstrate why they are not always applicable. A comprehensive survey about known detection methods is given by [17]. An updated discussion can be found in [18].

Firstly we need a method that can be reliably automated. A detection method for l_F should be based on reliable statistical tests and confidence levels. Any test that requires a visual inspection by the analyst is not acceptable for reasons of credibility and accuracy. This is important especially for automated output analysis. For example a popular method for detecting the transient phase is the method of [23]: the original output data is smoothed to distinguish between high a low frequency fluctuations. The analyst has to decide when the curve flattens out and, thus, selects the truncation point. Because of the necessary interaction of the analyst, this method has limited use in automated output analysis.

Practically all known detection methods are based on detection of constant mean values; see [10]. The crossing of the mean rule in [9] is one of the earliest recommended tests. It fails in the detection of l_F for simple output process with a constant mean but growing variance; e.g. for $X_i = ci\epsilon_i$, where c is a constant value and ϵ_i is a Gaussian white noise process. This process is unstable, however, the crossing of the mean rule will deliver an estimate l_E. A robust method should not return an estimate if the output process is unstable.

In [20] and [21] the detection of a truncation point is based on the assumption that the transient behaviour is a displacement of a stationary process: $X_i = \mu_i + X'_i$, where the sequence $\{\mu_i\}_{i=1}^{\infty}$ is converging and X'_i is stationary. The transient behaviour of the M/M/1 server, which is depicted in Figure 1, violates this assumption for all three different initialisations. Therefore, these detection methods cannot be recommended to detect l_F in queueing systems.

Several more recent methods are based on the variance of the sample mean. For example in [4] or [11] the variance of the sample mean during the transient phase is assumed to be greater then during steady state. This does not hold in general, and especially for the M/M/1 server, initialised empty and idle, this assumption may be too strict. In Figures 1(a) and 1(b) one can see that $\mathrm{Var}[R_i] < \mathrm{Var}[R_j]$ if $i < j$.

Another approach to reduce the initial bias is discussed in [2]. The described approach aims at estimating a truncation point after which random variables can be assumed to be identically distributed. Because it is aiming at equally distributed data, this approach is superior over other truncation point detection rules, which demand e.g. a constant mean only. It is applicable for one single simulation run as well as for multiple replications. The need for a fully automated approach by avoiding unspecified parameters is underlined. However, algorithmic properties of this approach, such as time complexity, storage requirements and sequential execution, are not discussed. This approach is discussed from the point of view of steady state analysis of mean values only.

A possible source of error is that the convergence is often assumed to be monotonic. This cannot be assumed in general. A process may even be periodic. In this case the process is not stable at all and l_F cannot be estimated. Methods which try to estimate a truncation point l_F by combining data of different observation indexes i are in danger of overlooking the instability. For example the method of Welch fails if the selected window size is similar to the length of the periodic cycle. This is demonstrated in [3]. The same problem occurs for the method of [18]. If the batch size is similar to the length of the periodic cycle this method fails.

The previous discussion on well known methods for reducing initial bias suggests that in automated analysis a detection method of l_F should require very little about the output process. Instead we propose a direct implementation of (1) in [6], [5] and [7]. The use of p replications executed in parallel enables establishing an independent and identically distributed random sample, consisting of p observations, of each X_i. These random samples can be compared with each other. A homogeneity test can be used to test equality in distribution. In [6] and [5] the Anderson-Darling k-sample test ([1] and [19]) is used for this purpose. This method processes the data by splitting it into a transient phase and an assumed steady state phase. The size of the transient phase is increased sequentially in order to find a valid truncation point. The size of the assumed steady state phase is increased proportionally so that the proportion between the size of the transient phase and assumed steady state phase is constantly $1 : r$. The random sample at the current candidate truncation point is compared with the random sample of the assumed steady state phase. If the homogeneity test supports the hypothesis of equality in distribution, a valid truncation point is assumed to have been found. Otherwise, this random sample is regarded as transient and more data is added to the assumed steady state phase. This method is able to deal with all the special cases described above (see [6], [5] and [7]). It can be used in analysis of arbitrary performance measures, like the mean, the variance or quantiles. As far as we are aware, this approach is the only method that aims at first order stationarity of the simulation output process.

The pseudo code of this approach is given in Listing 1. For convenience some special notation is used. Using the operators $+$, $-$, $/$ and $:=$ in conjunction with random variables X_k or S, see lines 3, 9, 12 and 13, means to use these operators on each component of the relevant sorted random sample separately. By S we denote a random sample $\{s_j\}_{j=1}^{p}$ of size p that is the sum of all ordered sequences which are not part of the transient period. Let $\{x_{jk}\}_{j=1}^{p}$ be the observed random sample of X_k and let $\{y_{jk}\}_{j=1}^{p}$ be the associated sorted sequence, then $s_j = \sum_{k=l+1}^{n} y_{jk}$. New observations are added, see line 9, whereas observations of the transient period are subtracted from S, see line 12. Dividing each component s_j by the number of addends results in an estimate of $F_{X_\infty}(x)$. The operator $\approx$ in line 15 and 19 refers to the homogeneity test. The procedure *autoSelect* sets r automatically according to the properties of the output process. This automatic selection starts with $r = r_{min}$ and increases r until a random sample is found, that is not equally distributed. If $r \geq r_{max}$ it can be assumed that there is no initial transient present. We used $r_{min} = 10$ to assure that the part of the assumed steady state phase is one magnitude larger than the assumed transient phase. We used $r_{max} = 10^5$ which is a valid choice for all our test models and many other commonly used simulation models. For details on the selection of r see [6]. The procedure *uniform(a,b)* delivers a uniform distributed integer random number between a and b used as index. This algorithmic approach is efficient because its run time complexity is $O(np \log(p))$, where p is the number of parallel replications and n is the number of collected observations of each replication. A proof of the time complexity and more details about the algorithmic approach can be found in [7].

Listing 1: Pseudo code of the algorithmic approach.

```
0  int n := 0;                 // 1 ≤ n < ∞
   int l := 0;                 // 1 ≤ l < n
   int r := autoSelect();      // r_min ≤ r ≤ r_max
   S := 0;  S' := 0;  // averaged random samples

5  bool NoTestFailed:=false;
   while (¬NoTestFailed){
     n := n + 1;
     observe(X_n);  // sample of p observations
     S := S + X_n;
10   if (0 ≠ n mod (r + 1)) continue;
     l := l + 1;
     S := S − X_l;
     S' := S/(n − l);
     NoTestFailed:=true;
15   if (¬(F_{X_l}(x) ≈ F_S(x))) NoTestFailed:=false;
     for (int k := 1; k ≤ r; k := k + 1){
       if (¬NoTestFailed) break;
       int u := uniform(lk + 1, l(k + 1));
       if (¬(F_{X_l}(x) ≈ F_{X_u}(x))) NoTestFailed:=false;
20   }
   }
```

5. TRUNCATION POINT DISTRIBUTION

Various estimators have various forms of probability distributions, e.g. means from normal populations are usually t-distributed. Here, we show that there is no typical probability distribution of l_F. The form of the distribution of l_F depends on the output process itself. With the following examples we demonstrate that the distribution of l_F can be very different. The empirical CDFs, which are depicted in this section, are based on more than 10^4 replications. We used 100 replications for each estimate of l_F and repeated the experiment more than 100 times. The algorithmic approach of Listing 1 combined with the Anderson-Darling k-sample test, as described by [5], is used in all experiments. For parametrisation the recommended standard values are used, i.e. we used 100 parallel replications and significance level $1 - \alpha = 0.95$.

The transient phase of the first example is governed by a quadratic displacement:

$$X_i = \begin{cases} \epsilon_i + \frac{k}{l^2}(i - l^2) & \text{if } i < l, \\ \epsilon_i & \text{else,} \end{cases}$$

where k is the offset. Here, X_i has a well defined truncation point $l = l_F$. $\mathrm{E}[X_i]$ is governed by a parabola for $i < l$. For all simulation experiments with this model we chose $k = 10$ and $l_F = 100$. In Figure 2(a) we can see that the estimates are distributed around observation index 90. The distribution is a step function that appears to have a symmetric density. Most estimates are a bit smaller than the optimal l_F as the convergence in (1) is from below. Note that approximate equality is implemented by the significance level of the homogeneity test, which is $1 - \alpha = 0.95$ for all experiments in this section.

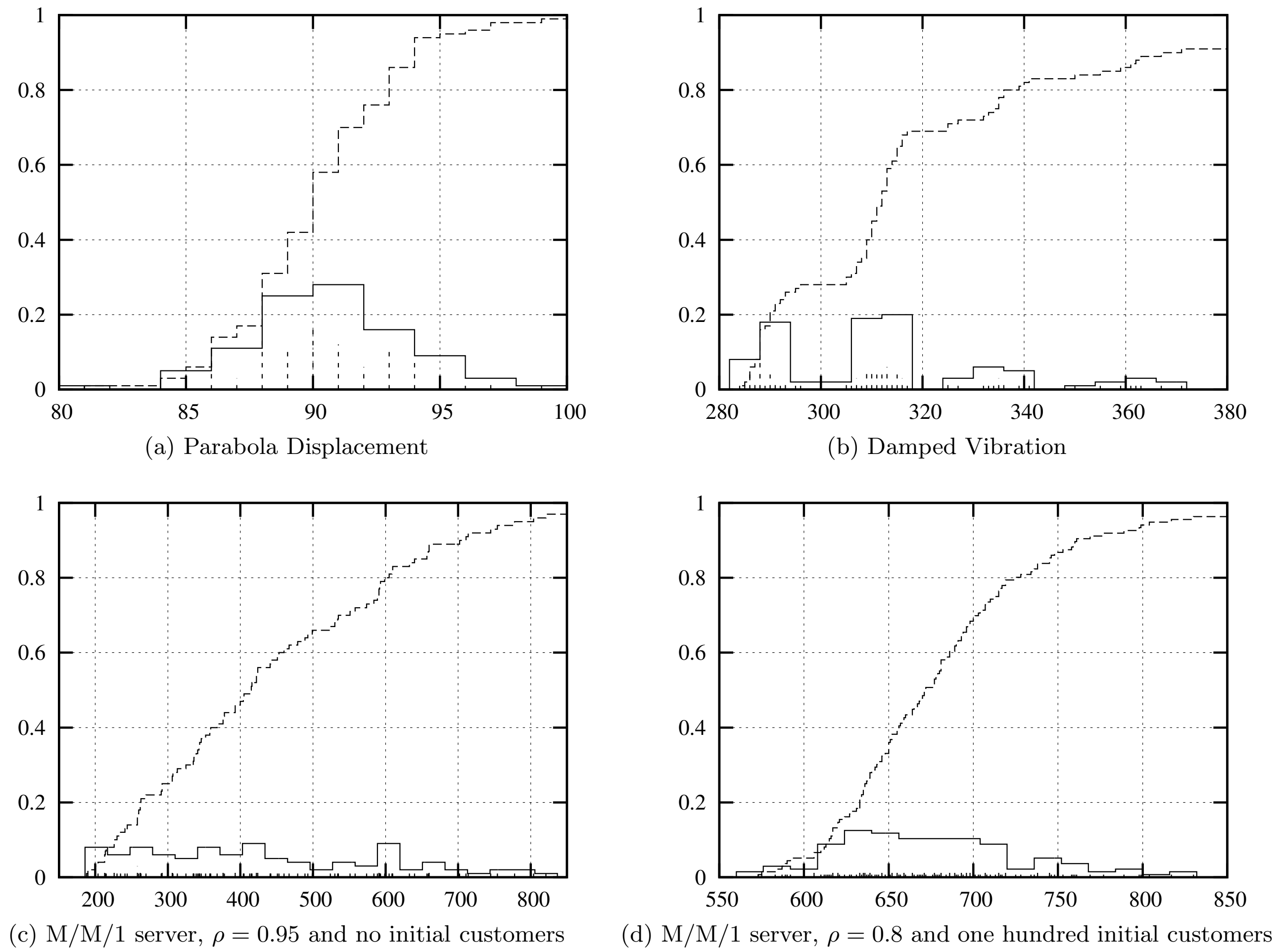
The next output process is a damped vibration.

$$X_i = \epsilon_i + (ke^{i\frac{\ln(0.05)}{l}}) \cdot \cos(i\omega),$$

where k is the amplitude and $T = \frac{2\pi}{\omega}$ is the cycle length. k is damped by an exponential function. At $i = l$ the exponential function is 0.05, therefore, a truncation point $l_F \geq l$ can be regarded as a suitable truncation point. In our experiments we used $k = 10$, $T = 50$ and $l = 250$. In Figure 2(b) one can see that all estimated l_F are greater than l. Because of its non-monotonic transient behaviour the distribution of the estimated truncation points is multi-modal. The maxima, resp. minima, of the amplitudes of the damped vibration are directly visible in the empirical distribution. Whenever the process is close to a maximum or minimum it is unlikely that the method detects a truncation point. The maxima and minima of our analysed output process are located at $\frac{1}{4}jT$, where j is a positive integer value. Compare this locations with the observation indexes 300, 325 and 350 in Figure 2(b).

In Figure 2(c) the distribution of the estimated truncation points of the M/M/1 server with $\rho = 0.95$ and no initial customer is depicted as our third example. Again, the observed measure is the systems response time R_i of the ith customer. The density distribution seems to have a long right tail and there is no clear mode. The convergence of $F_{R_i}(x)$ towards $F_{R_\infty}(x)$ is slow and the range of the distribution is large.

Figure 2(d) shows again the distribution of l_F for R_i of the M/M/1 server, but here $\rho = 0.8$ and there are one hundred initial customers. The density distribution is right-skewed. The range of the distribution is much smaller than that of the previous example. This shows that the high initial state



(a) Parabola Displacement

(b) Damped Vibration

(c) M/M/1 server, $\rho = 0.95$ and no initial customers

(d) M/M/1 server, $\rho = 0.8$ and one hundred initial customers

Figure 2: Histogram and empirical CDF of estimated truncation points.

introduces a transient behaviour which is easier to recognise.

The experiments in this section support the earlier statement that the probability distribution of l_F can have various forms. This depends on the output process itself. We can see that the estimation of l_F is robust and valid for a wide range of transient behaviours. In general we expect a symmetric density if l_F is well defined as in the first example. If $F_{R_i}(x) \rightarrow F_{R_\infty}(x)$ as it tends to infinity, we expect a density function of l_F with a longer right tail. Slow convergence is very challenging for any truncation point estimation method.

There are two possibilities when estimating l_F. On the one hand we may wish to estimate l_F as accurately as possible. This involves cases when l_F is possibly too small and some bias possibly remains in the assumed steady state phase. However, this approach assures that the deleted amount of data is small. This approach is recommendable if the collection of observations is expensive in sense of computer run time and is described in Listing 1. Alternatively, we may wish to be confident that the estimate of l_F is greater than its theoretical value. Although this might be more wasteful, however, bias is reduced. This approach can be integrated easily into the previous truncation point detection method by deleting all the observations collected so far, including a part of the steady state phase. This might be recommendable for our third example to assure that the truncation point is deep in the steady state phase.

6. CONCLUSIONS

We have shown that the rates of convergence to steady state in simulation are different for different performance measures. If steady state in terms of the probability distribution is detected, then all parameters specifying this probability distribution have achieved their steady state values. Subsequently, output data collected during this stage of simulation will produce estimates of arbitrary performance measures which is unbiased by the initial transient conditions. If steady state in terms of a specific moment is detected only, then such a claim is valid only for estimates of that moment. For example, the truncation point l_E can be use in mean value analysis, but it would be generally inappropriate to use it as the truncation point for steady state quantiles. However, in both cases, one could use the truncation point l_F.

Earlier proposed methods for determination of truncation points are confined to special classes of output processes and are applicable in mean value analysis. The method proposed in this paper is valid in analysis of arbitrary performance measures. It is based on the definition of the truncation point l_F, resulting from (1). Therefore, it is more robust than any method which implements only (2), (3) or both. An application of this method in analysis of multiple steady state quantiles is demonstrated in [8].

Empirical distributions of l_F obtained from our simulations show that the form of such a distribution can vary from almost symmetric to strongly right-skewed. The shape depends on the type of convergence of a given process to steady state (it could be monotonic, non-monotonic or even oscillating), as well as on the convergence rate.

Our method is based on homogeneity tests of independent and identically distributed random samples of different observation indexes. This is possible due to running multiple replications concurrently. We can see that output data analysis based on multiple parallel replications has got properties that make analysis of probability distributions feasible.

7. REFERENCES

[1] T. Anderson and D. Darling. A test of goodness of fit. *Journal of the American Statistical Association*, 49:765–769, 1954.

[2] H. P. Awad and P. W. Glynn. On an initial transient deletion rule with rigorous theoretical support. In *Proceedings of the 2006 Winter Simulation Conference*, pages 186–191, 2006.

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

[4] C. R. Cash, D. G. Dipplod, J. M. Long, W. P. Pollard, and B. L. Nelson. Evaluation of tests for initial-condition bias. In *Proceedings of the 1992 Winter Simulation Conference*, pages 577–585, 1992.

[5] M. Eickhoff. Steady state quantile estimation. In *Proceedings of the 13th GI/ITG Conference on Measurement, Modeling and Evaluation of Computer and Communication Systems (MMB'06)*, pages 155–171, 2006.

[6] M. Eickhoff. *Sequential Analysis of Quantiles and Probability Distributions by Replicated Simulations*. Ph.D. thesis, Computer Science and Software Engineering, University of Canterbury, 2007.

[7] M. Eickhoff, D. McNickle, and K. Pawlikowski. Efficient truncation point estimation for arbitrary performance measures. In *Proceedings of the 3rd Industrial Simulation Conference (ISC'2005)*, pages 5–12, 2005.

[8] M. Eickhoff, D. McNickle, and K. Pawlikowski. Using Parallel Replications for Sequential Estimation of Multiple Steady State Quantiles. In *Proceedings of the 2nd International Conference on Performance Evaluation Methodologies and Tools (Valuetools'07)*, 2007.

[9] G. S. Fishman. *Concepts and Methods in Discrete Event Digital Simulation*. John Wiley & Sons, Inc., 1973.

[10] A. V. Gafarian, C. J. Ancker, and T. Morisaku. Evaluation of commonly used rules for detecting steady state in computer simulation. *Naval Research Logistics Quarterly*, 25:511–529, 1978.

[11] D. Goldsman, L. W. Schruben, and J. J. Swain. Tests for transient means in simulated time series. *Naval Research Logistics*, 41:171–187, 1994.

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

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

[14] A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addision-Wesley, 1989.

[15] D. C. McNickle. Estimating the average delay of the first N customers in an M/Erlang/1 queue. *Asia-Pacific Journal of Operational Research*, 8:44–54, 1991.

[16] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1984.

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

[18] S. Robinson. A statistical process control approach to selecting a warm-up period for a discrete-event simulation. *European Journal of Operational Research*, 176(1):332–346, 2007.

[19] F. W. Scholz and M. A. Stephens. K-sample Anderson-Darling tests. *Journal of the American Statistical Association*, 82(399):918–924, September 1987.

[20] L. W. Schruben. Detecting initialization bias in simulation output. *Operations Research*, 30(3):569–590, May-June 1982.

[21] L. W. Schruben, H. Singh, and L. Tierney. Optimal tests for initialization bias in simulation output. *Operations Research*, 31(6):1167–1178, November-December 1983.

[22] K. S. Trivedi. *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. John Wiley & Sons, Inc., 2002.

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