

Remove Initial Bias From Steady-State Stochastic Simulation

Laura de Jesus dos Santos Varela Barroso
Instituto Superior Tecnico, Universidade de Lisboa
Lisboa, Portugal
laura.barroso@tecnico.ulisboa.pt

ABSTRACT

When a simulation begins, the queues are in a non steady-state phase and when the simulation is settled, they move towards the steady-state phase. It is known that the observations collected during the transient periods do not characterize the steady-state system and, consequently, it is important to discard the initial observations, known as initial bias. The main purpose of this project is to study the main topics in stochastic simulation output analysis, more specifically study a way to remove the initial bias.

Author Keywords

Stochastic simulation; initial bias; warm-up period; precision of simulation results; stopping rules for steady-state simulation

INTRODUCTION

In telecommunications, stochastic discrete event simulations have been vastly used to analyze and plan new communication networks as it offers flexibility of performance modeling, without the restriction of computer and time resources and costs. [8] Stochastic simulations are a powerful tool for exploring and predicting behavior of stochastic dynamic systems as telecommunication networks. In fact, stochastic discrete-event simulation has already become a commonly used tool of scientists and engineers in this area [1, 9].

Since random processes are simulated, any stochastic simulation needs to be seen as a (simulated) statistical experiment and, therefore, it needs to have several methods of analysis of (random) simulation output data.

To analyze the output data, there are many approaches that are used. Most of the used methods are focused on removing initial bias from a steady-state simulation, which means that is much needed to overcome the initial transient problem.

The urge to eliminate initial bias from steady-state simulation is a widely known challenge in the output simulation analysis area. The main reason for this situation to happen is because non-terminating simulations do not have pre-defined run lengths or initial conditions.

Several methods will be discussed and will be compared in detail to review their performance. This work's focus relies

on the analysis of simulation output with the purpose of removing the initial bias.

RELATED WORK

Stochastic Simulation

Based on a stochastic model, it is created a stochastic simulation that gives a projection which is based on a set of random values. This projection can be an estimation of a system's behavior since many real-world models of experiences in several areas are too complex to be analyzed with mathematical techniques and they are often time consuming, expensive, or merely impossible to execute.

The main advantage of using stochastic simulation is that it is possible to describe most complex, real-world systems with stochastic elements and evaluate them analytically. On the other hand, each run of a stochastic simulation can only translate into estimates of a system's real characteristics for a distinct set of input parameters.

Steady-State Simulation

Simulations can be analyzed as being either terminating or non-terminating. A terminating simulation runs for a certain duration of time T_E , where E is a given event (or a set of events) which concludes the simulation. Such simulation starts at time T_i under established initial conditions and stop at the time T_E .

Non-terminating simulation is a system that is running continuously in time. A steady-state simulation is a simulation whose purpose is to study the long-run behavior of a non-terminating system.

Being steady-state does not imply that the random variables will take on the same values every time but it does mean they will have nearly the same distribution [3, 4]. The rate of convergence of the transient distribution depends on the initial parameters and conditions. On the other hand steady-state distribution does not [9, 4].

Initial Bias

Telecommunications is an area in which dynamic systems with random data and results are becoming extremely common. Although there had been a great advance in the simulation techniques and methods, the results are merely random

samples. So, the studies about those results are not taken serious without a proper statistical analysis.

Once the observations collected through the transient periods do not characterize the steady-state, it is logical to discard the initial observations, known as *initial bias*, before going through an analysis of the results.

Graphical Methods

Welch's Method

Regarding the graphical methods that is the subject of this section, the Welch's Method is the simplest and most common graphical technique used for determining the warm-up length. This method requires several replications. The Welch's Method is explained in the following steps [4, 10]: 1.

Make n replications of the simulation where each has a run length m . Let Y_{ji} be the i^{th} observation from the j^{th} replication. Thus i takes values from 1 to m and j from 1 to n . 2.

Determine the ensemble averages over the replications. There will be \bar{Y}_i 's where:

$$\bar{Y}_i = \frac{1}{n} \sum_{j=1}^n Y_{ji} \quad \text{for } i = 1, 2, \dots, m \quad (1)$$

3. Define a moving average $\bar{Y}_i(w)$ to smooth the oscillations with high frequency in $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_m$. Where w is the window and is a positive integer. w is less than or equal to $m/4$. $\bar{Y}_i(w)$ is as follows:

$$\frac{1}{2i+1} \sum_{s=-(i-1)}^{i-1} \bar{Y}_{i+s} \quad \text{for } i = 1, 2, \dots, w \quad (2)$$

$$\frac{1}{2w+1} \sum_{s=-w}^w \bar{Y}_{i+s} \quad \text{for } i = w+1, w+2, \dots, m-w \quad (3)$$

4. Plot $\bar{Y}_i(w)$ for $i = 1, 2, \dots, m-w$ and choose L to be the value of i beyond which $\bar{Y}_i(w)$ seems to be converged.

SPC Method

Statistical Process Control (SPC) is another graphical method to analyze the output data of a stochastic simulation. The purpose of SPC is to determine when a process is in the presence of initial bias that affect the process, so the process may need to be corrected. So, a simulation that is in steady-state could be considered to be in control, while during a transient state it can be considered to be out of control. The method is now described in these steps [7]:

1. This method requires multiple replications. Let Y_{ij} denote the observation taken from each replication, where i is the number of the observation and j is the number of the replication. Then it is needed to calculate the ensemble averages. Ensemble averages $\bar{Y}_i : i = 1, 2, \dots, m$ are batched in b batches of size k . The means of the batches are represented as $\bar{\bar{Y}}_i, x = 1, 2, \dots, i$. The resulting time series after batching is represented as:

$$\bar{Y}_{(k)} = \bar{\bar{Y}}_1(k), \dots, \bar{\bar{Y}}_b(k) \quad (4)$$

2. A control chart must be constructed with the above time series. The estimates of population mean μ and standard deviation σ are then calculated from the last half of the series $Y(k)$. After that, the control limits are calculated with the formula:

$$CL = \hat{\mu} \pm z\hat{\sigma}\sqrt{b/2}, \quad z = 1, 2, 3 \quad (5)$$

Steady state is reached when the process is in-control and remains in-control. So we have to determine if the rules below are being followed:

- A point plots outside a 3- σ control limit.
- Two out of three consecutive points plot outside a 2- σ control limit.
- Four out of five consecutive points plot outside a 1- σ control limit.
- Eight consecutive points plot on one side of the mean.
- Initial points all plot to one side of the mean (as per expected bias).
- The warm-up period for the model can be selected by identifying the point at which the time-series data is in-control and remains in-control.

Statistical Methods

Kelton and Law Regression Method

Kelton and Law address to the problem of the initial bias with regression analysis. Their main assumption is that in a time-series of batch means, if the sequence of batch means that succeeds the initial transient can be approximated to a straight line with a zero slope, then the initial transient has ended after batch b_0 .

To fit the regression line, the method starts from the end of the series and then moves the sector backwards until the slope of

the line is different from zero. If there is not a zero slope in the line that is initially fitted, the replication length must be increased. However, the number of batches is kept fixed.

Randomization Tests for Initialization Bias

This method was present by Yucesan and it is based on the random shuffling of batched data and testing for the significance in the difference of two means. The primary advantage of this method is that no assumptions are required. These are steps that are needed to perform the method:

- The simulation need to run for a length of time m hours. Let Y_i be the i^{th} observation from the simulation output which is run for m hours.
- It is obtained an output time series Y_1, Y_2, \dots, Y_m .
- The data is put into b batches of length k .
- Then b batch means $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_b$. are obtained.
- The batch means are split into two groups. In the first iteration, the first group needs to include the first batch mean and, for the second group, it must contain remaining $b - 1$ batch means.
- In each iteration, the grand means of the two groups must be analyzed and compared. The statistic used test by Yucesan is the absolute value of the difference between the grand means of the two groups. If the difference is significantly different from zero between the two grand means, the null hypothesis described earlier is rejected. If the hypothesis made is abandoned, the groups must be rearranged; so the second batch is added to the first group and the second group will contain $b - 2$ batch means and the last step needs to be repeated. If it was accepted, then the group 2 data is the steady state simulation output.

Heuristic Methods

Conway Rule

Conway suggested in this method to truncate the initial data with the aim to reduce initial bias. This is executed in the initial runs to decide upon a stabilization period. It implements a series of measurements until the first of the series is neither the maximum nor the minimum of the remaining set of the simulation data. After that, the period is deleted from the result of each run. The method's algorithm is constructed using the steps:

1. It is needed to decide n and m the number of preliminary replications and the length of the preliminary replications, respectively.

2. Then it is computed y_{jr}^+ and y_{jr}^- with the following formulas:

$$y_{jr}^+ = \max(y_{jl} : l = r, \dots, m) \quad j = 1, \dots, n \quad (6)$$

$$y_{jr}^- = \min(y_{jl} : l = r, \dots, m) \quad j = 1, \dots, n \quad (7)$$

3. For $r = 1, 2, \dots, m$ it is needed to determine t_j such that $t_{j=\min} \{y_{jr}^- < y_{jr} < y_{jr}^+\}$ appears for the first time.

4. The truncation point t^* is then estimated and is given by $\max\{t_1, t_2, t_3, \dots, t_n\}$.

Crossing of the Means Rule

While the simulation output data is being generated, a running cumulative mean is being calculated. A count, represented by ω , is the result of the number of times the output data cross the cumulative mean. The higher the value of that count ω , the more confidence exists is the fact that the initial bias has been removed. The method's algorithm is based on the following steps:

1. The simulation output is generated $\{Y_1, Y_2, \dots, Y_m\}$.
2. The number of times the series crosses the mean is given by ω .
3. Then is made the calculation of $\Omega_1, \Omega_2, \dots, \Omega_j$ such that at l the number of crossings is equal to the pre-defined number.

M/M/1

The $M/M/1$ queuing model represents the queue length in a system having a single server and this system assumes a Poisson arrival process and both the arrival time and service time are exponentially distributed. An $M/M/1$ queue is considered to be a stochastic process.

- Let λ be the arrivals rate that reaches the system according to a Poisson process and moves the process from state i to $i + 1$.

- Let μ be the service times with an exponential distribution in the $M/M/1$ queue. The mean service time is determined with $1/\mu$.

This model with only one server serves its clients one at a time according to a FIFO discipline [5]. When the job is over, the client leaves the queue. The buffer has infinite size, so there is no limit regarding the number of requests it can accommodate.

This model is considered stable only if $\lambda < \mu$, which translates into the arrival rate λ must be less than the maximum rate service μ . If, on average, requests arrive faster than the job is done, then the queue will grow indefinitely long and the system will not have a stationary distribution.

Comparison between Matlab and other simulation softwares

When selecting a software to run the simulation, it needs to be flexible enough to run smoothly without erroneous results. There are a number of advantages when using a simulation package instead of a programming language:

- Simulation packages can provide many features that are needed when building a simulation model, so there is not a waste of time while programming and there is a cost reduction.
- Simulation models are easier to make modifications and maintain when written in a simulation package.
- It is easier to detect errors because simulation packages check automatically potential types of errors.

On the other hand, many programming languages are used to write simulation models:

- The programming language that is being used is well-known by modelers.
- A simulation model well-written in C, C++, or Java may require less execution time than a model developed in a simulation package. This happens because simulation packages are designed to address a great variety of systems with one set of modeling constructs.
- Software cost is generally lower.

ARCHITECTURE

The system used is the $M/M/1$ and is widely used in the field of simulation, and allow the study of the behavior of the stochastic simulations to be used.

Through the arrival rate and the service rate of this system, it is possible to calculate the most diverse answers. The system utilization factor, the percentage of time the server is occupied, is calculated by the expression $\rho = \lambda / \mu$.

These type of systems have theoretical formulas for various parameters, by which become an important means of comparison during the experimental run.

A representation of the system is represented in Figure 1:

Implementation of Methods

In this work, it will be used the $M/M/1$ system as a base for this study, starting with the arrival rates of $\lambda = 50$, $\lambda = 80$, $\lambda = 90$, and $\lambda = 95$. It is important to note that these values are base values and may be adapted to each method.

Results of interest and performance measures

There are some results of interest that can also represent some system evaluation measures amongst others:

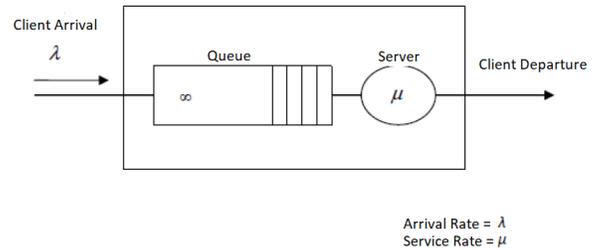


Figure 1. $M/M/1$ system representation

- Time in system
- Percentage of server usage
- Number of clients in the system

All of these responses allow to evaluate the system's performance. Once the results of interest have been selected, it is important to choose which measures of performance that will be used on the collected data from these same results.

IMPLEMENTATION

The process of implementing the work is described. Also, the collection and treatment of the results that will be used in this work will be carried out.

$M/M/1$ System

After the simulation of the $M/M/1$ system, the result is a set of files, where each of these files corresponds to a value ρ (utilization factor) that is going to be studied within the system. Each of these files contains the time in the system registered for the different clients at each replication. Other parameters like the number of replications and the number of clients will contribute to this set of files:

- The number of replications was set to 20. Except when the method does not need any replications.
- The number of clients is intended to be 500000, although it has high computational costs. Other tested values are 50000 and 5000.

Welch's Method

The steps and equations for calculating Welch's Method are listed below:

- A number of replications $rep = 20$ is performed, each of length win , where win is the window size and it is much larger than the anticipated truncation point. The observations are averaged over all replications at each time-step to create the average process.

- The moving averages are plotted for several values of widow size, *win* .
- The arrival rate is defined by λ .
- If no value of w is satisfactory, the number of replications must be increased.
- The truncation point is selected visually from the moving averages plot.

SPC Method

This method implements a control hart, which with an upper and lower control limit quantifies the bounds of natural variation. This implementation determines the mean and variance based on the mean observations over the replications.

- The number of replications *rep*, the arrival rate λ and the number of clients *cli* was defined previously while running a *M/M/1* simulation.
- Based on the given parameters and the given rules, it is possible to visually understand when the system is "*under control*" according to theoretical prepositions of this method.

Kelton and Law Regression Method

Kelton and Law use the least square regression procedure for a first degree polynomial and determine its slope. The steps for implementing this method are listed below:

- The function receives as parameters a matrix *tsys* with a matrix of the number of replications and the number of observations (clients), and also receives the window size, which must be equal to the number of observations. The function returns the slope of the function in each coordinate.
- The parameter *tsys* results from an execution of an *M/M/1* simulation with the number of replications *rep*, the arrival rate λ and the number of clients *cli* defined previously.
- The returned value *ans*, the closer it is to zero, the greater that the confirmation the initial transient has ended with the sequence of batch means that succeeds the initial transient can be approximated to a straight line with a zero slope.

Randomization Tests for Initialization Bias

The implementation of this function is as follows:

- This method requires 6 parameters: *yucesan* (*y*, *b*, *m*, *corr*, *NS*, *NL*) where *y* is the vector of the observations collected (the dimension must be at least *b*), *b* is the number of initial lots, *mis* the initial size of the lots, *corr* is the value of the maximum correlation allowed between lots, *NS* is the number of shuffles, *NL* is the desired level of significance.
- The result is the number of observations that must be present in each batch, as well as the number of observations that must be removed. The value of the variable *corr* is used as a threshold.

Conway Rule

In this method it will be followed the suggestion of 20 replications (it can vary around that value) and the number of clients will vary from 5000 to 500000. This rule implements two measurements until the first of the series is neither the maximum nor the minimum of the remaining set of the simulation data.

Crossing of the Means Rule

This method determines the number of mean crosses, and stops when those crosses reach a pre-defined threshold and returns the observation number. It is recommended the threshold to be defined at 25 for the an high arrival rate, and the threshold to be defined at 7 for the a low arrival rate, so this parameters will vary around these values and it will be adapted to each situation.

DESIGN

It is in this chapter that the design of the solution will be carried out. The adjustment of the parameters in each method is needed to understand which method is the most adequate for each situation.

Design process in Welch's Method

Before the adjustment of each method, it is necessary to do a previous study of this method's behaviour with certain pre-defined parameters.

There is an output of an *M/M/1* simulation with this method applied with 50.000 clients, 500 size window, 20 replications and arrival rate of 90%. If we "zoom in" in the x-axis of the plot, it is possible to observe more clearly where the bias ends, as it is possible to see in Figure 2.

Applying the method with several arrival rates, maintaining 20 replications, it is possible to observe the initial bias, and select how many clients to discard. The results are in Table 1.

This method has gained wide acceptance due to its simplicity, assumptions and ease of implementation. Its accuracy presented also to be good. But, since the method is based on the use of cumulative statistics, it may be a little conservative,

overestimating the warm-up period. Observing the practical values, with and without bias, it can be verified that the system behaves near-expected, and it may be possible to state that the initial bias no longer contaminates the system values.

λ	"Bias"	Theo. t_{sys}	Pract. t_{sys}	Pract. t_{sys} w/o Bias
50	20	0.02	0.02002	0.02021
80	70	0.05	0.04745	0.05119
90	170	0.1	0.1034	0.1019
95	550	0.2	0.1846	0.2072

Table 1. Table of Discarded Bias on $M/M/1$ with Welch's Method

Design process in SPC Method

For a warm-up of this method's study, the first simulations had parameters such as the arrival rate set at 50 and the number of clients set to 5000, in order to gain "sensitivity" to the functioning of this method. The second simulations had as parameters the arrival rate set at 80 and the number of clients set at 50000, with the purpose to study a system with high traffic intensity. If we "zoom-in" in the same graphic, it can be seen the initial bias before reaching 100 clients in Figure 3.

Applying the method with many arrival rates, also maintaining 20 replications, it is possible to observe the initial bias, and select how many clients to discard. The results are in Table 2:

λ	"Bias"	Theo. t_{sys}	Pract. t_{sys}	Pract. t_{sys} w/o Bias
50	220	0.02	0.0198	0.02057
80	300	0.05	0.0589	0.05249
90	2700	0.1	0.0949	0.0989
95	3100	0.2	0.2342	0.1922

Table 2. Table of Discarded Bias on $M/M/1$ with SPC's Method

With the SPC method to set upper and lower limits, it can be seen graphically that there are clients whose system time (t_{sys}) exceeds the limits and, therefore, the rules imposed by the method are not always fulfilled. On the other hand, comparing the theoretical values with the practical values without

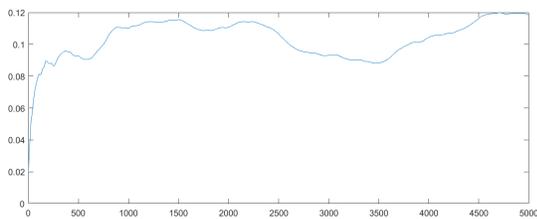


Figure 2. "Zoom-in" Welch's Method output in $M/M/1$ system simulation with $\lambda=90$

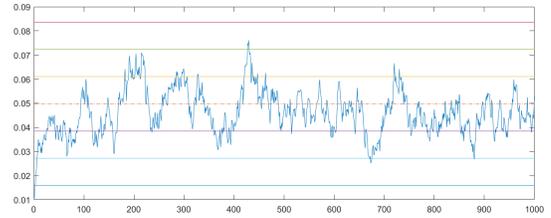


Figure 3. SPC Method output in $M/M/1$ system simulation with $\lambda=80$

initial bias, it can be seen that the method performed well, especially in $M/M/1$ with arrival rates of 90% and 95%.

Design process in Kelton and Law Regression Method

This method is about observing if the slope at each batch is close to zero. For each arrival rate, we have a table of vector to see which ones are closer to zero. It is important to emphasize that the number of slopes in the vector varies according to the number of clients divided by the size of each batch and is therefore variable and adapted to each situation.

For $\lambda=50$, the regression method is applied in every batch. Also, for this arrival rate, we have the slopes vector:

[-0.0012 ; 0.0004 ; 0.0031 ; 0.0014 ; 0.0024 ; -0.0013 ; -0.0009 ; 0.0005 ; 0.0025 ; 0.0011]

For $\lambda=80$, the regression method is applied in every batch. For this arrival rate, we have the slopes vector with less number of values due to the window adjustment to see the initial bias:

[-0.0206 ; -0.0050 ; -0.0090 ; -0.0027 ; 0.0011]

For $\lambda=90$, the regression method is applied in every batch. Once again, for this arrival rate, we have the slopes vector with less number of values due to the window adjustment, and, also, due to the adjustment in the number of clients for a better visualization

[0.0680 ; 0.0109 ; -0.0370 ; -0.0421 ; 0.0028 ; -0.0091 ; 0.0040 ; -0.0053]

Finally, for $\lambda=95$, the regression method is applied in every batch and it is represented in Figure 4.

Once again, for this arrival rate, we have the slopes vector with less number of values due to the window adjustment, and, also, due to the adjustment in the number of clients for a better visualization of the initial bias:

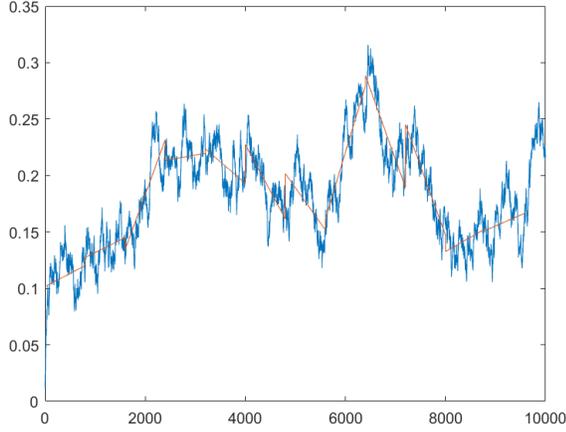


Figure 4. Kelton and Law Regression Method output in $M/M/1$ system simulation with $\lambda=95$

[0.0184 ; 0.0174 ; 0.0971 ; 0.0063 ; -0.0305 ; -0.0659 ; -0.0499 ; 0.1177 ; -0.0992 ; -0.0980 ; 0.0203 ; 0.0155]

Applying this method with several arrival rates, considering that the batch size is relevant, it is possible to observe the initial bias, and select how many batches (of variable size number of clients) can be discarded. The results are in Table 3. It is important to remind that the window size can change to allow to see more clearly the warm-up phase, so the results of the initial bias in the Table 3 are generalized.

λ	"Bias"	T. <i>tsys</i>	P. <i>tsys</i>	P. <i>tsys</i> w/o Bias
50	1 (± 500)	0.02	0.02245	0.02035
80	1 (± 1000)	0.05	0.0578	0.05144
90	1 (± 1200)	0.1	0.1125	0.1032
95	2 (± 2000)	0.2	0.2438	0.2036

Table 3. Table of Discarded Bias on $M/M/1$ with Kelton and Law Regression Method

In terms of the elimination of the initial bias and the estimation of the warm-up period, this method had a good performance in most of the cases. It may be concluded that this method works well to remove the initial bias, although it is more general and less detailed, unless there is a larger study on the size of the batches, which results in a vector of slopes with greater number of values to analyze, which makes the analysis of this method more complex.

Design process in Randomization tests for initialization bias

The input of this function is as follows:

`yucesan (y, b, m, corr, NS, NL)`

In this input, the y parameter is the vector of the observations collected (the dimension must be at least b) and it is collected

by running a $M/M/1$ simulation with a defined number of clients and an arrival rate value, b is the number of initial lots, m is the initial lot size, $corr$ is the maximum correlation value allowed between lots and the value suggested was 0.05 or 0.1, NS is the number of shuffles which the suggested value was the maximum possible, for example 9999, and, finally, NL is the desired level of significance, which was recommended to be 0.1.

This method presents some problems when providing the initial parameters, which may be difficult to choose in the beginning. In table 4 are presented the input parameters and the initial bias that this methods recognizes.

λ	b size	m size	Number of clients	Initial bias
50	50	50	50000	750
80	50	50	50000	1100
90	15	15	500000	1700
95	10	10	500000	2800

Table 4. Table of results on $M/M/1$ with Yucsan Method for several λ values

Applying this method with several arrival rates, considering that the batch size is relevant, it is possible to observe the initial bias, and select how many batches (of variable size number of clients) can be discarded and we can observe the theoretical and practical bias. The results are in Table 5:

λ	"Bias"	T. <i>tsys</i>	P. <i>tsys</i>	P. <i>tsys</i> w/o Bias
50	750	0.02	0.02525	0.02110
80	1100	0.05	0.0628	0.04970
90	1700	0.1	0.1415	0.1018
95	2800	0.2	0.2551	0.1880

Table 5. Table of Discarded Bias on $M/M/1$ with Kelton and Law Regression Method

In relation to the observations that should not be used when analyzing the results, the initial bias, it can be possible to verify that its value increases as the value of the arrival rate increases. This allows to conclude that, as we increase the utilization factor of the system, the longer it takes to stabilize, that is, to enter its stationary state.

It is clear that this method discards more initial bias than others, like in Welch's method. This method's results showed that it did not always give particularly safe estimates of the length of the initial transient phase, especially in a system with low traffic intensity. It is possible to conclude that, in those cases, there were some overestimation of the warm-up phase.

Design process in Crossing of the Means Rule

This method determines the number of mean crosses, and stops when those crosses reach a pre-defined threshold and

returns the observation number. The challenge of this method is to determine which is the most adequate threshold for each situation. This method does not need any replications.

The author of this method, Fishman, recommends the threshold to be defined at 25 for the an high arrival rate, and the threshold to be defined at 7 for the a low arrival rate, so this parameters will vary around these values and it will be adapted to each situation.

For $\lambda=50$, the number of crosses observed are in Table 6.

	50	35	25	15	5
5000 clients	159	118	88	43	9

Table 6. Table of counts on $M/M/1$ with Crossing of the Means Rule for $\lambda=50$

For $\lambda=80$, the number of crosses observed are in Table 7.

	75	50	35	25	15	5
50000 clients	625	460	306	192	54	15

Table 7. Table of counts on $M/M/1$ with Crossing of the Means Rule for $\lambda=80$

For $\lambda=90$, the number of crosses observed are in Table 8.

	75	50	35	25	15	5
500000 clients	847	628	449	255	83	16

Table 8. Table of counts on $M/M/1$ with Crossing of the Means Rule for $\lambda=90$

For $\lambda=95$, the number of crosses observed are in Table 9.

	75	50	35	25	15	5
500000 clients	847	628	449	255	83	16

Table 9. Table of counts on $M/M/1$ with Crossing of the Means Rule for $\lambda=95$

The conclusions to be drawn from this method are made differently from the rest, because here we are not explicitly told how much "data" to remove, so that the system is free of the initial bias, but it is said that the warm-up phase ends, or rather, the confidence that the warm-up phase has ended is higher as the higher the cross number becomes (within reliable intervals), taking into account the threshold set for a certain arrival rate.

Design process in Conway Rule

The Conway Rule is described as: the warm-up period is the first point in the series that is neither the maximum nor the minimum of the remaining observations.

In the Table 10, for each arrival rate, it is shown where the truncation point is, which allows us to understand where the warm-up phase ends.

λ	"Bias"	T. <i>tsys</i>	P. <i>tsys</i>	P. <i>tsys</i> w/o Bias
50	4976	0.02	0,02127	0.02466
80	4988	0.05	0,04983	0.05387
90	4801	0.1	0,09299	0.11091
95	4991	0.2	0,15449	0.15908

Table 10. Table of Discarded Bias on $M/M/1$ with Conway Rule

Of the results that can be observed in the table above, it can be seen that the truncation point is much above that the previous methods point out. This may indicate that there is overestimation in situations of low traffic intensity and high traffic intensity as to where the warm-up phase actually ends, which decreases the accuracy of this method.

This method indicated an extremely high truncation point for the considered arrival rates, which leads to the conclusion that it is not a suitable method for the removal of the initial bias.

CONCLUSIONS

This document describes of a number of methods for detecting and removing the initial bias. They were divided into three groups: graphical methods, heuristic methods and statistical methods. The advantages and disadvantages of each method have been discussed. This investigation of the initial transient phase detection problem included a comparative study of two of methods in the three mentioned groups . Those methods were implemented and their performance investigated through the use of simulation studies in the $M/M/1$ queuing system.

After observing the results and comparing the practical values with the theoretical values, from which it was possible to discuss results, used parameters and conclusions obtained, here the final conclusions will be drawn around the six methods that were studied.

Welch's Method vs. SPC Method

Although the SPC method has very strict rules in terms of acceptance of the fact that the system is "in control" in visual terms, Welch's method is much more clear and explicit in relation to the initial bias visualization in the $M/M/1$. This system is quite simple, so the visual aid to identify the initial

bias is the main source of the conclusions. Also, the practical results, in comparison with the theoretical values, were not inappropriate, but rather viable. As we can observe, as the arrival rate increases, also the number of discarded data increases but nothing extraordinary, maintaining a coherence that the theoretical information about the method had confirmed. Welch's method revealed to be useful to the purpose, without overestimation of the truncation point.

Welch's method does require the estimation of a number of parameters: run length, number of replications and specifically the window size. But, the parameters can only be estimated by trial and error, which contributes to the user gaining sensitivity to the topic of the initial bias, which has a lot of contrast with the SPC method that, despite the rules being clear, comes with very restricted and rigorous assumptions.

The only objection which can be relevant is when automation the identification of the initial bias. Both methods requires the user to judge the smoothness and flatness of a moving average plot.

Kelton and Law Regression Method vs. Randomization Tests for Initialization Bias

Although these two methods fall into the category of "Statistical Methods", they are quite different. In terms of output, the rules defined by the "Kelton and Law Regression Method" give much more support in discarding the batches which contain the initial bias.

For instance, the Kelton and Law Regression Method is considered to be complex to code and the large number of parameters that require estimation ends up by withdrawing the focus on the identification of the bias itself. Also, the shuffling of the batches of data requires a significant volume of calculations, making the execution of the procedure slow.

Conway Rule vs. Crossing of the Means Rule

Given the observed results, the Crossing of the Means Rule was the one that presented the best performance, even with the subjectivity associated to the threshold defined before the simulation, but chosen according to the author's suggestion. The method performed as expected and mentioned previously in [6, 2], in which the number of crosses increased, thus giving more confidence that the initial bias had in fact ended.

For example, observing the results of Conway Rule, there was a great overestimating in those results, over the remaining other methods. This may be due to the fact that the method involves two rules that restrict very much the choice of how far the warm-up phase goes, and is, therefore, a rather meticulous way to chose the the truncation point that has diminished the performance of this method.

So, within the category of "Heuristic Methods", the Crossing of the Means Rules was the one which accomplished best in removing the initial bias, given in to consideration the less "good performance" presented by the Conway Rule.

Future Work

Graphical methods revealed to be useful as the user involvement in the estimation of the warm-up period can be seen as advantageous. The heuristic methods have the advantage of providing specific and rigorous rules for determining the warm-up period, making automation of the procedure possible.

As a suggestion, it would be beneficial to incorporate one or more methods in a hybrid framework with a chosen truncation method. A example would be to incorporate Welch's Method with Kelton and Law Regression Method, both with visual aid of the system's behaviour but with very different rules when determining of how much data must be discarded.

It is still a challenge to determine the accuracy and generality of the different approaches, and decide which fits best the intention of the user. Most of the literature search, and most of the evidences suggests that the most commonly used methods are the simple time-series inspection or Welch's method, both of which have their shortcomings, although it was proven that Welch's Method had a good performance.

Finally, the development of new methods should continue. The aim should be to develop methods that are effective in terms of the criteria set out, like simplicity, ease of implementation of the method, accuracy, assumptions needed, and parameter estimation.

A number of methods for determining the warm-up period or detecting the initial bias have been described under three headings: graphical methods, heuristic methods, and statistical methods. The advantages and disadvantages of each method have been discussed. It is clear that no single method can be recommended, although there are better suggestions than other. It is still recommended that a number of methods are employed if an accurate estimate of the warm-up period is required for future projects.

REFERENCES

1. Barradas, R. Construo de metamodelos de simulao estocstica utilizando splines, Oct. 2015.
2. Gafarian, A. V., Anker, C. J., and Morisaku, T. Evaluation of commonly used rules for detecting steady state in computer simulation. In *Proceedings of the 32Nd Conference on Winter Simulation, WSC '00*, Naval Research Logistics Quarterly (1978), 511–529.
3. Grassmann, W. Initial bias and estimation error in discrete event simulation. In *Proceedings of the 14th*

- Conference on Winter Simulation - Volume 2*, WSC '82, Winter Simulation Conference (1982), 377–384.
4. Law, A. *Simulation Modeling and Analysis*, fifth ed. McGraw-Hill, New York, NY, USA, 2015.
 5. Law, A., and Kelton, W. *Simulation Modeling and Analysis*, third ed. McGraw-Hill, New York, NY, USA, 2000.
 6. Mahajan, P., and Ingalls, R. Evaluation of methods used to detect warm-up period in steady state simulation. In *Proceedings of the 36th Conference on Winter Simulation*, WSC '04, Winter Simulation Conference (2004), 663–671.
 7. Robinson, S. New simulation output analysis techniques: A statistical process control approach for estimating the warm-up period. In *Proceedings of the 34th Conference on Winter Simulation: Exploring New Frontiers*, WSC '02, Winter Simulation Conference (2002), 439–446.
 8. Santos, M. Nonlinear regression metamodels: a systematic approach. *International Journal of Simulation & Process Modelling* 5, 3 (2009), 241–255.
 9. Taylor, S. Analyzing methods of mitigating initialization bias in transportation simulation models, 2010.
 10. Welch, P. The statistical analysis of simulation results. In *Computer Performance Modeling Handbook*, S. S. Lavenberg, Ed. Academic Press, New York, NY, USA, 1983, 268–328.