Evaluation of initialization bias in simulation output, comparing Schruben-based methods

Elisabetta Roviera
Politecnico di Torino
s328422

***Abstract*—Initialization bias in simulation models arises from the arbitrary selection of initial conditions, resulting in systematic distortions in early outputs. This bias can severely compromise steady-state simulations, as early transients may not accurately reflect the system's long-run behavior, potentially leading to misguided conclusions and faulty decision-making. Detecting and mitigating this bias is therefore critical, especially in disciplines such as operations research, manufacturing, and supply chain optimization.**

To address this challenge, a range of statistical tests have been devised to assess whether a simulation has reached steady state. In this study, I examine three key methods:

- **Detect Initialization Bias Method – This approach analyzes the autocorrelation structure of the simulation data to uncover transient effects.**
- **Test Initialization Bias Method – This method employs statistical hypothesis testing to assess whether early observations deviate significantly from the expected steady-state mean.**
- **Transient Mean Tests (BM, Area, Maximum, BM+Area, BM+Maximum) – These tests evaluate various metrics of transient behavior, offering a comprehensive view of initialization bias.**

These tests were implemented and applied to two benchmark examples from the paper, thereby validating their accuracy and reliability. Additionally, a final example featuring a carefully initialized simulation model was analyzed. In this final example, an absence of bias was consistently indicated across all tests, demonstrating their effectiveness in confirming a steady state. Overall, these findings underscore the importance of robust bias detection to ensure that simulation outputs remain reliable and free from transient distortions.

This is the report of the notebook file "InitializationBias", which contains a detailed explanation of all algorithms, the Python implementation of these algorithms, two examples from the paper for each algorithm to verify their functionality, and the final example.

I. Detecting Initialization Bias in Simulation Output

Initialization bias distorts steady-state simulation results due to non-representative initial conditions. This method provides a statistically rigorous approach to detect, not correct, this bias in the average simulation output. It transforms the output into a standardized test sequence, separating the deterministic bias "signal" from the stochastic "noise". This sequence is then compared to a Brownian bridge process.

The test examines the maximum deviation of standardized partial sums from zero. A significant deviation, particularly an early maximum, suggests bias. A statistical value is calculated from this deviation and compared to critical values from an F-distribution for hypothesis testing. Under the null hypothesis of no bias, the sequence should behave like a Brownian bridge.

Algorithm 1 Algorithm for Detecting Initialization Bias

1: **Find the Global Maximum.** Determine $\hat{s}$, the global maximum of the normalized partial sums:

$$\left\{ \frac{k\, S_n(k)}{\sqrt{n}} \right\}, \quad k = 1, \dots, n,$$

and record its location $\hat{k}$.

2: **Estimate the Variance.** Compute the variance $\hat{\sigma}^2$ using the residual variance from an autoregressive model of order p (AR(p)) fitted to the latter part of the output:

$$\hat{\sigma}^2 = \hat{\sigma}_\epsilon^2 \left(1 - \hat{\phi}_1 - \hat{\phi}_2 - \cdots - \hat{\phi}_p\right)^{-2}.$$

3: **Compute the Degrees of Freedom.** Estimate the degrees of freedom ν as follows:

$$\nu = \frac{n\, \hat{\sigma}_\epsilon}{2\, \hat{\sigma} \left(p - \sum_{i=1}^{p} (p - 2i) \hat{\phi}_i\right)}.$$

4: **Calculate the Test Statistic.** Set $\hat{t} = \hat{k}/n$ and compute

$$\hat{h} = \frac{\hat{s}^2}{3\, \hat{\sigma}^2\, \hat{t}\, (1 - \hat{t})}.$$

5: **Determine the Significance Level.** Compute the p-value $\hat{\alpha}$ using the upper tail of the F-distribution:

$$\hat{\alpha} = \bar{F}_{3,\nu}(\hat{h}),$$

where $\bar{F}_{3,\nu}(\cdot)$ denotes the upper tail probability for an F-variates with 3 and ν degrees of freedom.

6: **Decision Rule.** Reject the null hypothesis of no negative bias if $\hat{\alpha} < \alpha$, where α is the predefined significance level.

II. Initialization Bias Detection Method

Initialization bias, i.e., bias due to non-representative initial conditions in a simulation, can distort steady-state performance estimates. A statistical test is proposed to detect such bias by transforming the simulation output into a standardized sequence that separates the deterministic signal (potential bias) from the stochastic noise. In the absence of bias, this standardized sequence should behave like a Brownian bridge; significant deviations, particularly during the early stages, indicate the presence of initialization bias.

Algorithm 2 Algorithm for Detection of Initialization Bias

1: **Data Preprocessing and Stationarity Determination:**

2: **(a)** Estimate the effective sample size n_{eff} using an AR(p) model if a stationary start index is not available.

3: **(b)** Define the stationary segment as

$$\{Y_{n_{\text{eff}}+1}, \dots, Y_n\}.$$

4: **(c)** Compute the stationary mean:

$$\bar{Y}_{\text{stationary}} = \frac{1}{n - n_{\text{eff}}} \sum_{i=n_{\text{eff}}+1}^{n} Y_i,$$

and the corresponding standard deviation $\sigma_{\text{stationary}}$ (using an unbiased estimate, ddof=1).

5: **Residual Calculation and Modified Brownian Bridge Construction:**

6: **(a)** For each observation, calculate the residuals:

$$\text{residuals}_i = Y_i - \bar{Y}_{\text{stationary}}, \quad i = 1, \dots, n.$$

7: **(b)** Compute the cumulative sum of the residuals:

$$S(k) = \sum_{i=1}^{k} \text{residuals}_i, \quad S(0) = 0, \quad k = 0, 1, \dots, n.$$

8: **(c)** Construct the modified Brownian bridge:

$$B(k) = \frac{S(k) - \frac{k}{n} S(n)}{\sigma_{\text{stationary}} \sqrt{n_{\text{eff}}}}, \quad k = 0, 1, \dots, n.$$

9: **Test Statistic Calculation:**

10: **(a)** Compute the overall mean:

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i,$$

and the sample standard deviation:

$$\sigma_{\text{sample}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \bar{Y}_n\right)^2}.$$

11: **(b)** For each $k = 1, \dots, n$, define $\bar{Y}_k$ as the mean of the first k observations.

12: **(c)** Calculate the test statistic:

$$T = \frac{\sqrt{45}}{n^{3/2}\, \sigma_{\text{sample}}} \sum_{k=1}^{n} \left(1 - \frac{k}{n}\right) k \left(\bar{Y}_n - \bar{Y}_k\right).$$

13: **Critical Value Evaluation and Decision Rule:**

14: **(a)** Obtain the critical value from the t-distribution:

$$t_{\text{critical}} = t.\text{ppf}(1 - \alpha, \text{df} = n_{\text{eff}}),$$

where α is the significance level.

15: **(b)** If $T > t_{\text{critical}}$ (or with the sign reversed for a test of positive bias), reject the null hypothesis of no bias; otherwise, do not reject.

III. Transient Mean Detection Method

Initialization bias in simulation output poses a significant problem as transient means can distort the estimation of steady-state performance. To address this, a method is proposed that aims to detect these transient means, ultimately helping to identify potential initialization bias within simulation data. The central concept of this method involves a comparison of variance estimates derived from the initial and final segments of a simulation run. The underlying premise is that if the simulation is in a steady state (the null hypothesis), the variance estimates from these two segments should be statistically similar. This method is built upon the assumption that the simulation output behaves as a stochastic process, characterized by a transient mean that gradually diminishes towards a steady state and exhibits serial correlation in the data. To facilitate this variance comparison, the method employs several variance estimators, including Batch Means, Area-based estimators utilizing Standardized Time Series, and Maximum Deviation estimators. The actual detection of transient means is performed using an F-test, which is designed to compare the variance estimates from the early portion, denoted as $V_{k,b'}$, and the late portion, denoted as $V_{k,b-b'}$, of the simulation. The test statistic, calculated as $F = \frac{V_{k,b'}}{V_{k,b-b'}}$, is then evaluated against a critical value from the F-distribution. Should the F-statistic exceed this critical value, the null hypothesis of stationarity is rejected, leading to the conclusion that transient means are present and suggesting the potential existence of initialization bias.

Algorithm 3 Algorithm for Detecting Transient Means

1: **Divide the Simulation Output.** Partition the time series into two contiguous, non-overlapping segments, selecting the number of batches b' and $b - b'$.

2: **Estimate the Variance.** Compute variance estimates $V_{k,b'}$ and $V_{k,b-b'}$ using one of the following methods:

- *Batch Means:* Compute variance from batch means across segments.
- *Standardized Time Series (Area Method):* Use the integral of the standardized series over time.
- *Maximum Deviation:* Compare the maximum deviations observed in the segments.

3: **Compute the F-Statistic.** Compute

$$F = \frac{V_{k,b'}}{V_{k,b-b'}}.$$

4: **Determine the Significance Level.** Compare the computed F value with the critical value $f_{d_1,d_2,\alpha}$ from the F-distribution with degrees of freedom d_1 and d_2 corresponding to the number of batches in each segment.

5: **Decision Rule.** If $F > f_{d_1,d_2,\alpha}$, reject the null hypothesis and conclude that transient means are present in the simulation output.

IV. Final Example

This example applies the three initialization bias detection algorithms to a simulation model designed to represent a real-world inventory system. The model simulates the evolution of the **average stock level**, relevant to contexts such as warehouse management or logistics operations.

Two distinct scenarios are compared within the simulation:

1) **Biased Initialization Scenario:** In this case, the simulation starts with a low initial stock level. The stock level then experiences a growth phase, modeled by a logistic function with added noise. This growth phase introduces a clear initialization bias. Subsequently, the system converges to a stationary behavior around a predefined target stock value.
2) **Unbiased Initialization Scenario:** In contrast, this scenario begins with the system already in a stationary state. There is no initial growth phase, representing an ”unbiased” case from the outset.

The primary objective of this example is to analyze and compare the impact of initialization bias on the system's dynamics. Furthermore, it aims to evaluate the effectiveness of three implemented bias detection methods in identifying initialization bias in both the biased and unbiased scenarios. These methods are: **detect_initialisation_bias**, **test_initialisation_bias**, and **transient_mean_test**. The results from applying these algorithms to both scenarios will demonstrate their ability to correctly identify the presence or absence of initialization bias.

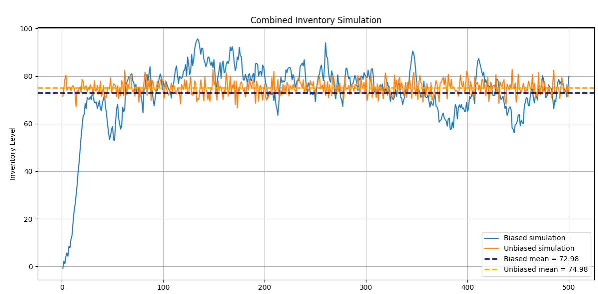

Fig. 1. Combined Inventory Simulation

a) Combined Inventory Simulation: The Combined Inventory Simulation graph in Figure 1 illustrates the simulated progression of the inventory level of a product in two scenarios: one characterised by an initial growth phase (blue curve) and one devoid of such a phase (orange curve). The blue dotted line denotes the average inventory level in the former scenario, while the orange line indicates the latter. Evidently, the initial growth phase exerts an influence on the average inventory value, which is lower than in the scenario devoid of such a phase. This demonstrates that initiating a system in a transitional phase introduces a bias into the estimation of the average inventory level.

b) Biased Simulation: The ACF plot shows a slow decay in autocorrelation with increasing lag, indicating that past values continue to influence future observations—a hallmark of initialization bias. Similarly, the PACF plot displays a pronounced spike at the first lag followed by a rapid drop within the confidence interval, suggesting that the dependency is predominantly short-term. These patterns are typical of biased data, where an initial artificial trend creates spurious correlations that fade over time. Statistical tests reinforce this observation: most methods, including *detect_initialization_bias*, *test_initialization_bias*, and several *transient_mean_test* variants (BM, Maximum, BM+Area, BM+Maximum), yield test statistics that exceed their critical values, leading to rejection of the null hypothesis. Although the *transient_mean_test - Area* method does not reject the null hypothesis—possibly due to lower sensitivity—the overall evidence strongly supports the presence of initialization bias in the dataset.

c) Unbiased Simulation: In contrast, the ACF plot for the unbiased simulation shows autocorrelation values that remain near zero across all lags, indicating little to no influence from past observations. The PACF plot corroborates this by showing no significant spikes beyond the first lag, with values staying within the confidence interval. Correspondingly, all applied statistical tests fail to reject the null hypothesis, as the *detect_initialisation_bias* and *test_initialisation_bias* methods, along with the various *transient_mean_tests* (BM, Area, Maximum, BM+Area, BM+Maximum), report statistics below the critical values. These results consistently suggest the absence of initialization bias, confirming that the tests appropriately indicate a lack of transient effects in the data.

V. Conclusion

In summary, this report presents an evaluation of three statistical methods— **detect_initialisation_bias**, **test_initialisation_bias**, and **transient_mean_tests**—for their effectiveness in identifying initialization bias in simulation output. Through benchmark examples and a final inventory simulation, the capability of these methods to discern between biased and unbiased simulation runs has been demonstrated. The tests generally aligned in their findings, successfully detecting bias in scenarios designed to exhibit it and confirming the absence of bias when expected. Notably, the inventory simulation highlighted the subtle nature of initialization bias and the importance of robust detection methods. While the transient_mean_test - Area method showed a slight inconsistency, the overall suite of tests provides a valuable toolkit for simulation practitioners. These methods offer a robust framework for assessing the reliability of simulation results by ensuring that outputs are not contaminated by transient effects, thereby enhancing the accuracy and credibility of simulation-based decision-making.

References

[1] A. Author, B. Author, *Title of the Paper on Initialization Bias*, Journal Name, Year.

[2] C. Author, D. Author, *Statistical Methods for Simulation Analysis*, Journal Name, Year.

[3] E. Author, F. Author, *Transient Behavior in Simulation Models*, Conference Name, Year.