

A note on the number of replication runs in stochastic traffic simulation models

Wilco Burghout

Introduction

While nowadays many articles on the application, calibration and validation of (microscopic) traffic simulation models discuss the need for a certain number of replications to obtain reliable results, none of them seems to discuss clearly how exactly to determine the number required.

Stochastic simulation

Most microscopic and many mesoscopic and macroscopic (dynamic) traffic simulation models have the property to represent random variations in the behaviour of the simulated traffic. In microscopic models this randomness, or stochasticity is used in many of the simulated processes. Examples are the arrival process of vehicles at the entries to the simulated network, the distribution of driver characteristics such as desired speeds, critical gaps and route choice preferences. Also decisions are simulated using random draws from certain distributions, for instance compliance to regulations, route choice and lane choice decisions.

These sources of randomness are used to represent the diversity of these characteristics, as they are found in reality. However, they have a large effect on the results of the simulation, and different simulation runs can therefore produce very different results, due to a different sequence of random numbers that happen to be drawn for the mentioned processes.

Replications

This stochasticity in the results of simulation runs has been widely recognized, but unfortunately they are sometimes ignored when such simulation models are applied. The usual way to deal with the random variation of the simulation results, given the same input data, is to replicate the simulation runs a number of times, and take the average of the measures of interest (such as travel times on links or routes, flows over a number of locations, speeds, etc.). These measures of interest are usually called Measures of Effectiveness (MOEs).

However, the number of replications needed is usually not discussed in any detail, and a number between 5 and 10 is usually taken. In some cases a formula is shown ((Toledo, T. *et al.* 2003), (Bourrel, E. 2003), (Ahmed, K. 1999)) but it is unclear where this formula comes from, or how it should be applied. Here I will try to start from the basics outlined in general simulation literature (Law, A.M. & Kelton, W.D. 2000) and discuss the implications for (micro) simulation of traffic.

Depending on the specific application of the simulation model, certain output quantities are defined to be of interest. These *measures of effectiveness* (MOEs) are usually flows and speeds over specific points (detectors), and travel times. For instance in case of a calibration study, one might want to calibrate the parameters of the model, in order to match flows and speeds over certain detector stations for which field data is available. When these MOEs have been defined, it is important that the simulated quantities can be compared to the field data. In order to do so, a good estimate of the mean value of these quantities across the simulation runs (which produce varying results) needs to be obtained.

In (Toledo, T. *et al.* 2003) and (Ahmed, K. 1999) the following formula for the required number of replications (N), given m initial replications was presented (rewritten with corrected notation):

$$N(m) = \left(\frac{S(m)t_{m-1,1-\alpha/2}}{\bar{X}(m)\mathbf{e}} \right)^2 \quad (1)$$

Where,

- $N(m)$ = The number of replications required, given m replications
 $\bar{X}(m)$ = The estimate of the real mean μ from m simulation runs (samples)
 $S(m)$ = The estimate of the real standard deviation σ from m simulation runs
 α = Level of significance
 ϵ = Allowable percentage error of the estimate $\bar{X}(m)$.
 $\epsilon = | \bar{X}(m) - \mu | / | \mu |$.
 $t_{m-1, 1-\alpha/2}$ = critical value of the two-tailed t-distribution at a level α of significance, given $m-1$ degrees of freedom.

Using this formula the required number of replications $N(m)$ can be calculated, based on an initial number of replications to obtain a good point estimate $\bar{X}(n)$ of μ , the real mean of the simulated quantity.

This formula is a derivation of a more general procedure described (Law, A.M. & Kelton, W.D. 2000). In order to obtain an estimate of the mean (across simulation runs) of a quantity of interest, a confidence interval for the mean is constructed. From statistics we know that the (t-)confidence interval for the mean μ for a small number of sample values can be constructed using the Student distribution (t-distribution):

$$\bar{X}(n) \pm t_{n-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{n}} \quad (2)$$

Where

- $\bar{X}(n)$ = The estimate of μ from n simulation runs (samples)
 $S^2(n)$ = The estimate of σ from n simulation runs (samples)
 n = Number of simulation runs
 α = Level of significance
 $t_{n-1, 1-\alpha/2}$ = Critical value of the t-test for $n-1$ degrees of freedom and significance α

The use of the t-distribution for the construction of the confidence interval rests on the assumption that the quantity under scrutiny is IID normal distributed over the simulation runs. While this assumption is seldom true, Law & Kelton show that the law of large numbers (central limit theorem) starts to work (in practice) already at $n=10$, and for larger n even correlation between the outcomes of different simulation runs does not affect the coverage of the confidence interval too much.

In formula (1) some additional simplifications are used to be able to calculate a fixed number $N(m)$ based on initial estimates of μ and σ , namely that the initial estimates $\bar{X}(m)$ and $S(m)$ based on m observations are close enough to the real mean and standard deviation, and thus do not change much when the number of replications is increased to $N(m)$.

The problem here is that given the (empirical) fact that the outcomes from simulation runs are seldomly normal distributed, and often correlated to some degree, the initial estimates $\bar{X}(m)$ and $S(m)$ are not very good, leading to an inaccurate prediction of the required number of replications.

Based on these considerations an incremental approach is presented (taken from (Law, A.M. & Kelton, W.D. 2000)), in which the estimates $\bar{X}(n)$ and $S^2(n)$ are recalculated when more replications are made.

1. Make an initial number of $m \geq 2$ runs and calculate initial (crude) estimates $\bar{X}(m)$ and $S^2(m)$.
2. Decide the size of the allowable percentage error $\epsilon = |\bar{X}(m) - \mu| / |\mu|$
3. Calculate the adjusted percentage error $\epsilon' = \epsilon / (1 + \epsilon)$
4. Decide the level of significance α
5. Calculate the new $\bar{X}(n)$ and $S^2(n)$
6. Calculate the half-length of the confidence interval: $\mathbf{d}(n, \alpha) = t_{n-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{n}}$
7. if $\frac{\mathbf{d}(n, \alpha)}{|\bar{X}(n)|} \leq \epsilon'$ use $\bar{X}(n)$ as an unbiased point estimate for μ , else make one more replication and go back to 5.

In this algorithm first initial estimates are generated in step 1, and the allowable percentage error is set in step 2. This error is adjusted in step 3 so that the *resulting* percentage error for $\bar{X}(n)$ is correct.

The advantage of this procedure is that the estimates $\bar{X}(n)$ and $S^2(n)$ for the mean and variance are becoming better with each iteration, as does the estimation of the resulting confidence interval. This procedure still assumes IID outcomes, but they need not be normal distributed.

Automation of replications

Anyone who has applied a (micro)simulation program knows that the number of quantities to be analysed can be very large. In (Toledo, T. *et al.* 2003) for instance, the flows and speeds at many locations were analysed, over many periods of time. Also the travel times for certain stretches in the network were analysed for different periods of time.

The procedure outlined above should be performed for *every* quantity (such as flow at a certain location) and *every* time period of aggregation. To provide an example, imagine a small application, where the data of 10 simulated detector stations are collected, and we only look at flows and speeds for the period of 3 hours (typical period of study). If we decide to take 15 minute averages, this means that there are 12 time periods, and each time period 10 average speeds and 10 flows are produced. This means that for each replication 240 confidence intervals have to be calculated. And since each replication changes the (initially rough) estimations for the mean and standard deviations, *each replication* these 240 confidence intervals need to be re-calculated. The replications continue until all 240 confidence intervals are small enough (according to the predefined allowed size)

It seems obvious that this procedure is impossibly time consuming to do by hand, and therefore I propose that this procedure is included as a standard component in traffic simulation models. The user just defines the Measures of Effectiveness, the allowable size of the confidence intervals (as a percentage of the mean) and the level of significance (alpha) for the t-tests that will be applied.

When using existing simulation packages, a shell program could be constructed that starts the simulation replications automatically, calculates the confidence intervals from the results and continues to make more replications until the required level of error has been reached.

This seems to be the most correct way to obtain reliable simulation results.

Simplifications

Until this procedure is automated, the best practice could be to have an initial number of replications to decide the confidence intervals of all (or a limited number of the) quantities, and only re-evaluate the quantities with the largest confidence interval during the replications. In addition the re-evaluation can be done every k (for instance 2 or 3) replications.

References

Ahmed, K., 1999. Modeling Drivers' Acceleration and Lane Changing Behavior, PhD thesis, ITS Program, Massachusetts Institute of Technology, Cambridge, MA,

Bourrel, E., 2003. Modélisation Dynamique De L'écoulement Du Traffic Routier: Du Macroscopique Au Microscopique, PhD Thesis thesis, L' Institut National des Sciences Appliquées de Lyon,

Law, A.M. & W.D. Kelton, 2000. Simulation Modeling and Analysis. McGraw-Hill,

Toledo, T., H. Koutsopoulos, A. Davol, M.E. Ben-Akiva, W. Burghout, I. Andreasson, T. Johansson & C. Lundin, 2003. Calibration and Validation of Microscopic Traffic Simulation Tools: Stockholm Case Study. Transportation Research Record (1831): 65-75.