Improving Probabilistic Traffic Modelling through Advanced Sampling

S.C. Calvert * [1]

H. Taale [2]

M. Snelder [1]

S.P. Hoogendoorn

All with the Transport & Planning Deptarment,

Delft University of Technology, Delft, The Netherlands

[1] Also with TNO, Delft, The Netherlands

[2] Also with Rijkswaterstaat & TrafficQuest, The Netherlands

* Email: s.c.calvert@tudelft.nl

1 Introduction

In traffic models certain presumptions are often made to simplify the complex systems that rule the world of traffic flow. This is necessary as not every variable can be considered. Furthermore, it is commonplace that equilibrium states are sought that give a good average, or rather deterministic, representation of the dynamics of traffic. Such an approach makes presumptions of traffic demand and supply for a (non-existent) average situation. However there must be a realisation that traffic is hardly ever ‘average’ [1]. It is especially in the terms ‘average’ and ‘deterministic’ that a realisation must exist that these terms are composed of extensively varying situations. By considering real stochasticity in these processes, a more complete picture of the traffic system is gained [2, 3].

2 Stochasticity in Traffic Models

Stochasticity is generally incorporated in traffic modelling through two main methods: analytical, or replicative simulations through Monte Carlo simulation or a derivative thereof. The latter is often seen as a simpler approach, however demands a high computational effort [4]. It is in an effort to reduce this computational load and speed up stochastic calculation, that advanced methods of sampling are investigated in this contribution for their ability to do so. This contribution describes two advanced sampling methods to reduce computational load in probabilistic traffic modelling, demonstrates their efficiency in an experimental case, and makes a comparison between the methods and simple Monte Carlo simulation. The objective of this contribution is to demonstrate the effectiveness of these methods for computational reduction when considering multiple variations in demand and supply variables. Initial results are shown hereof.

3 Advanced Sampling

The considered sampling methods in this research are Importance sampling and Latin Hypercube sampling. Simulations are also applied for simple random sampling as a reference method. Importance sampling is a technique used in Monte Carlo simulation that gives extra consideration to the outlying sections of a distribution which have a lower probability of being sampled, but have a relatively large influence on the output variable [5]. By giving the extremities of a distribution a greater probability than they originally, a higher chance of being sampled is created and therefore the rate at which the output distribution is ‘complete’ is greater.

Latin hypercube sampling is a stratified sampling technique that, other than general stratified sampling, ensures that the entire sample space for multiple input variables is sufficiently covered [6, 7]. The method is an extension of quota sampling. The basic method sees variables evenly sampled from the sample spaces, also known as a d-dimensional hypercube. Combinations of the samples are randomly generated, such that a good spread of samples is achieved to form a single target function. This can be applied on any number of dimensions of variables, but is applied in this research in two dimensions.

4 Case Study

A comparison is made between the two sampling methods and on a reference method. This is performed through their application in a dynamic macroscopic traffic model on a real urban freeway network. For each method the values of both the capacity and of the traffic demand are varied according to predefined distributions (see Fig. 1 and 2), which apply a certain multiplication factor for each simulation. Each method is applied with varying samples for 200 simulations. The travel time along the main corridor is recorded in each simulation. These results are collected and are analysed for convergence using distribution convergence using the squared error of the distribution (SEd) from the Central Limit Theorem (CLT). Furthermore an analysis is made of the resulting travel time distributions to test the accuracy of the results.

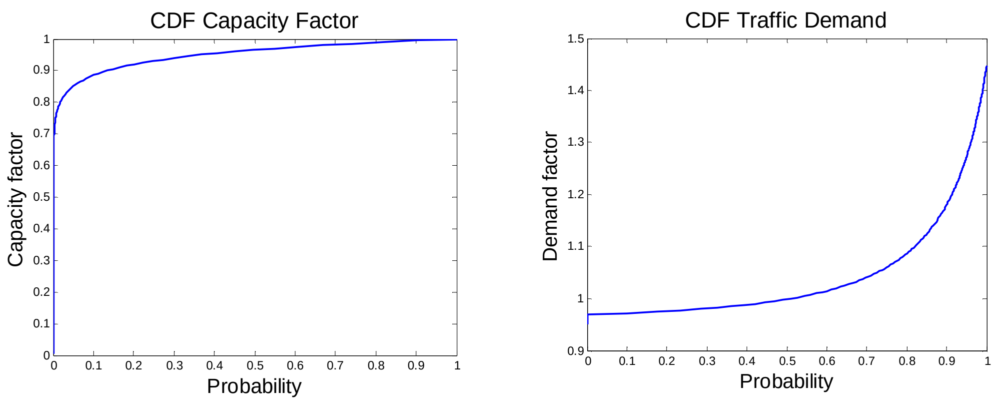

Fig. 1-2: Cumulative Density Function of the Capacity factor (left) and Traffic Demand factor (right)

The convergence method of the Central Limit Theorem (CLT) states that for large number of samples n, the random sum Sn has a distribution that increasingly approximates a normal distribution with expectation nμ and variance σ^2. The rate at which the distribution of Sn approaches the normal distribution acts as a convergence indicator tested against the squared error in the area of the distribution in comparison to the normal distribution. The CLT states that:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow Y \sim N(0,1) \quad as \;\; n \rightarrow \infty \tag{1}$$

Therefore the mean square error of the distribution is written as such:

$$e = \sum_i (Y_n - N(0,1))^2 \tag{2}$$

where: Y_n is the normalised distribution of S_n

The results of the three methods are collected as distributions from the CLT method. An example of the convergence towards the normal distribution for the CLT is given for the Latin Hypercube method for increasing iterations in Figure 3. Here the convergence towards the normal distribution is evident.

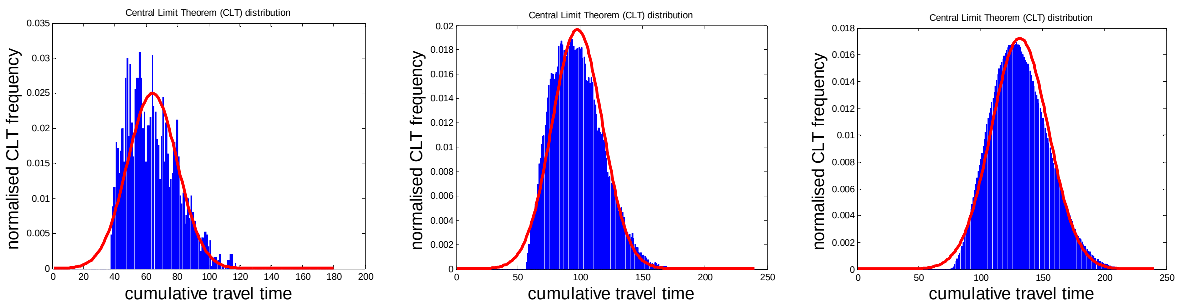

Fig. 3: Central Limit Theorem distributions for convergence. Travel time with input samples from Latin Hypercube.

5 Results

The CLT squared distribution error for increasing sample size n, depicting the rate of convergence, is given in Figure 4 for each method. There it is shown, on a logarithmic and linear scale that the Latin Hypercube method converges at a greater rate than that of simple random sampling. This is especially the case for the initial samples. However the improvement is not large. Importance sampling in this experiment however does not show a significant improvement, but rather a decrease in convergence speed.

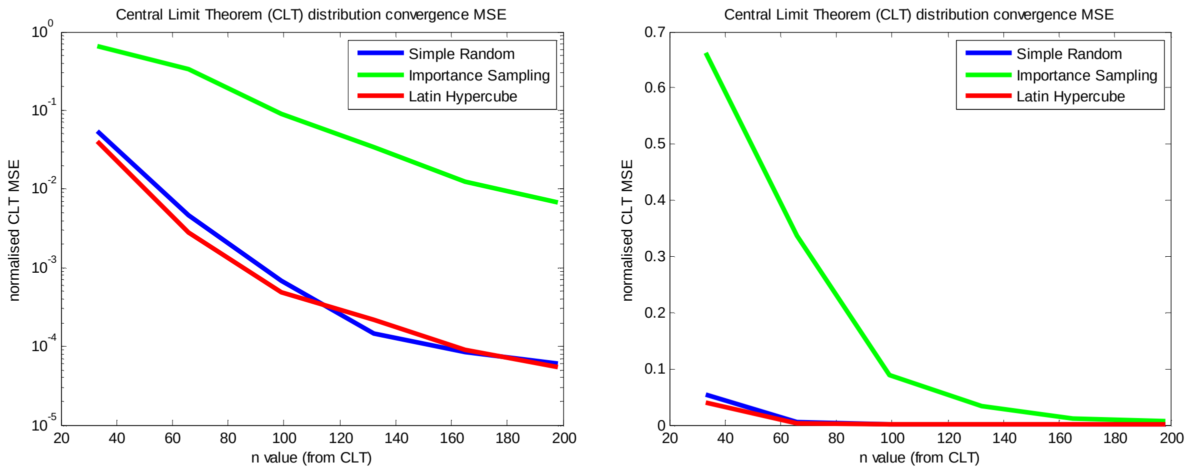

Fig. 4: Central Limit theorem convergence rate for three sampling methods on a logarithmic-scale (left) and on a linear scale (right).

A greater rate of convergence allows a practitioner to reduce the required number of simulations and therefore the required calculation time in applying the Monte Carlo technique. This is the case as seen with the improvement for the Latin Hypercube method. This result is in line with expectations, as the method is far more structured in the manner that samples are taken and therefore can give a far more representative distribution of the system with fewer samples than with random sampling. The method of Importance sampling should, in theory, also show an improvement. However this method has the drawback that a predefined ‘dummy’ distribution is required which acts to focus sampling on certain values. Determining the best distribution to apply is not trivial and may often require a certain amount of trial-and-error. In this research, up to now, the applied distribution is obviously not optimal, even after a few attempts. The fact that one requires a nominal ‘dummy’ distribution, however, does make the application of this method somewhat more cumbersome in comparison to methods that do not require such tweaking in advance.

6 Conclusion

In conclusion, the results of this research have so far shown that Latin Hypercube sampling is an effective method for reducing the computational load in such a way. Importance sampling was not shown to improve convergence, however this method is very dependent on the initial nominal distribution and therefore is harder to apply effectively. However the full effectiveness of Latin Hypercube method has not yet been conclusively proven. In the full paper further experimental results demonstrating the full potential of advanced sampling are given. A more comprehensive description is given of the methods and their mathematical properties. Convergence is also tested according to the Berry-Esseen theorem of CLT convergence. Furthermore a greater in-depth analysis is performed on all the results and their implementations for assisting probabilistic modelling.

References

[1] J. M. Ernst, C. M. Day, J. V. Krigmeier, and D. M. Bullock, "Probe Data Sampling Guidelines for Characterizing Arterial Travel Time," in 91th meeting of the Transportation Research Board, Washington DC, 2012.

[2] S. Clark and D. Watling, "Modelling network travel time reliability under stochastic demand," Transportation Research Part B: Methodological, vol. 39, pp. 119-140, 2005.

[3] J. W. C. van Lint, O. Miete, H. Taale, and S. P. Hoogendoorn, "A systematic framework for the assessment of traffic measures and policies on the reliability of traffic operations and travel time," 91th meeting of the Transportation Research Board, Washington DC, 2012.

[4] C. Chang, Y. Tung, and J. Yang, "Monte Carlo Simulation for Correlated Variables with Marginal Distributions," Journal of Hydraulic Engineering, vol. 120, pp. 313-331, 1994.

[5] D. P. Kroese, T. Taimre, and Z. I. Botev, Handbook of Monte Carlo Methods vol. 706: Wiley, 2011.

[6] R. L. Iman and W. Conover, "Small sample sensitivity analysis techniques for computer models. with an application to risk assessment," Communications in statistics-theory and methods, vol. 9, pp. 1749-1842, 1980.

[7] M. D. McKay, R. J. Beckman, and W. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," Technometrics, pp. 239-245, 1979.