USTM Resiliency Sensitivity Analysis

Gregory Macfarlane[a,*], Natalie Gray[a]

[a] *Civil and Environmental Engineering Department, 430 Engineering Building, Provo, Utah 84602*

Abstract

The input parameters used in travel demand models contribute to result uncertainty. A coefficient of variation was used to determine the range of possible values for each parameter. Sampling methods can approximate a normal distribution of parameter values with a discrete number of draws. This paper looks at Monte Carlo sampling and Latin Hypercube sampling. A three-step travel demand model is created with a 25-zone dummy model to evaluate if Latin Hypercube sampling reduces the number of draws needed to approximate Monte Carlo sampling. The mean modechoice logsum value was used to evaluate a cumulative standard deviation of values based upon 100 and 600 draws of parameter values for each sampling method. The standard deviation for Latin hypercube samples stabilize between 100 and 200 draws, whereas Monte Carlo samples often haven't stabilized at 600 draws. Latin hypercube sampling does reduce the number of draws needed, to where it can be applied to a large-scale model.

Keywords: Sensitivity Analysis, Resiliency, Latin Hypercube Sampling

1. Questions

There exists uncertainty in travel demand models. This is known by transportation planners but the majority do not use any particular method to quantify it. This uncertainty exists to some extent by the variance among input parameters. Two popular sampling methods to draw from the range of possible parameters are Monte Carlo (MC) simulation and Latin Hypercube Sampling (LHS). MC simulation requires large computations to be effective on a statewide model. LHS reduces the number of variants needed, but the amount of reduction is unknown. (Yang et al., 2013)

The research questions are therefore:

- Using a dummy travel demand model, can Latin Hypercube Sampling reduce the iterations needed to approximate random sampling methods (e.g., Monte Carlo simulation)?
- Does this method of sampling have few enough iterations for statewide model application?

*Corresponding Author

Email addresses: gregmacfarlane@byu.edu (Gregory Macfarlane), nat.gray2000@gmail.com (Natalie Gray)

2. Methods

To examine the effects of parameter input sensitivity, we developed a trip-based travel model with three steps:

1. trip generation,
2. trip distribution, and
3. mode choice.

Trip generation, the first step, was conducted using socioeconomic (SE) data from a 25-zone dummy model of the ActivitySim GitHub repository. Trip production was estimated using the 2017 National Household Travel Survey data (NHTS2017). The trip productions were summarized by household sizes, vehicles, and workers, and the weighted mean of each trip purpose was taken. The three trip purposes used are Home Based Work (HBW), Home Based Other (HBO), and Non-Home Based (NHB). NHTS2017 data and the SE data were merged based upon their household size, vehicles, and workers with maximum thresholds set as 4 persons, 3 vehicles, and 2 workers per household. Trip attraction was skipped for this analysis.

The second step, trip distribution, used distance and travel time skims from an example in the ActivitySim GitHub repository. The skims were simplified to use auto, nonmotorized, and transit modes. Travel time for auto used the single occupancy vehicle AM time, nonmotorized travel time used the walking distance skim multiplied by a factor of average walking speed, and transit time used the walk to local bus time.

Mode choice, the third step, calculates utilities for the three modes. These utilities were exponentiated, added together, and the natural log was taken to get a logsum value for every origin and destination pair. The utility equations for the mode choice model are as follows:

$$\text{drive\_utility} = (\text{coeff\_ivtt} * \text{auto}) + (\text{coeff\_cost} * \text{auto\_cost} * \text{DIST}) \tag{1}$$

$$\text{nonmo\_utility} = (\text{k\_nmot} + 20 * (\text{coeff\_walk1} * \text{nonmotor})) \tag{2}$$

$$\text{trans\_utility} = \text{k\_trn} + (\text{coeff\_ivtt} * \text{transit}) \tag{3}$$

The mode choice parameters (constants and coefficients) were obtained from the USTM Resiliency Model. These values are shown in Table 1 and Table 2.

With this simple three-step model, MC and LHS methods were used to determine the possible combinations of parameter variance. To identify a standard deviation for each parameter, a coefficient of variation was used. A set coefficient of variation of 0.30 was used for all six input parameters (Zhao and Kockelman, 2002). The standard deviation was equal to 0.30 multiplied by the mean, where the mean values in this situation are the base scenario parameters (as identified in Table 1 and Table 2).

Table 1: Mode Choice Coefficients

Name	HBW	HBO	NHB
CIVTT	-0.0450	-0.0350	-0.0400
CCOST	-0.0016	-0.0016	-0.0016
CWALK1	-0.0900	-0.0700	-0.0800
AUTOCOST	18.3000	18.3000	18.3000

Table 2: Mode Choice Constants

Name	HBW	HBO	NHB
K_TRN	-0.5140	-0.9853	-1.3020
K_NMOT	1.7602	0.5448	-0.5359

The MC random sampling uses the R function of `rnorm`. LHS uses the `lhs` package in R. Since this package only chooses variables on a zero to one scale, the values given use a function to put the random sampling on the right scale needed for the given parameter. The full code for both methods can be found in a public GitHub repository. 100 and 600 draws of random samples for both methods are generated. With these generated parameters, the mode choice model step was run for every set of input parameters for each purpose. The mean logsum value for each run was determined to compare each continuous draw.

3. Findings

The parameters generated were compared for both sampling methods. Figure 1 shows the distributions for the HBW parameters when using 100 draws, and Figure 2 shows how that changes when using 600 draws. These distributions show that LHS gives normally distributed parameters with fewer draws than MC sampling. At 100 draws LHS shows a nearly perfect normal distribution, where there are some discrepancies for the MC generated parameters. Without looking at the mode choice results, these Figures show that LHS is likely to estimate the full variance of the results with much fewer draws.

To determine if LHS is effective at a reasonable amount of iterations, the standard deviation was calculated for each additional draw. This value shows how much the mean mode choice logsum value for the 25 zones can vary. When the standard deviation for the draws stabilizes, that shows that the amount of generated parameters has captured all of the possible variances of the results. This can be visualized for each purpose. The HBW results for the cumulative standard deviation are shown in 3. The results for the other two purposes (HBO and NHB) are in 4 and 5 respectively.

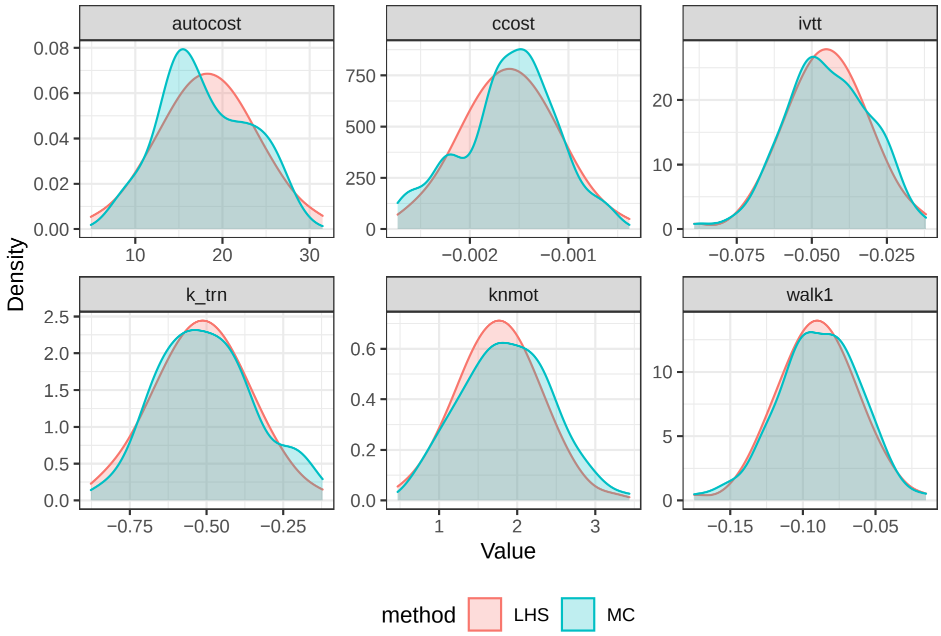

Figure 1: HBW Distributions for Input Parameters with 100 Draws

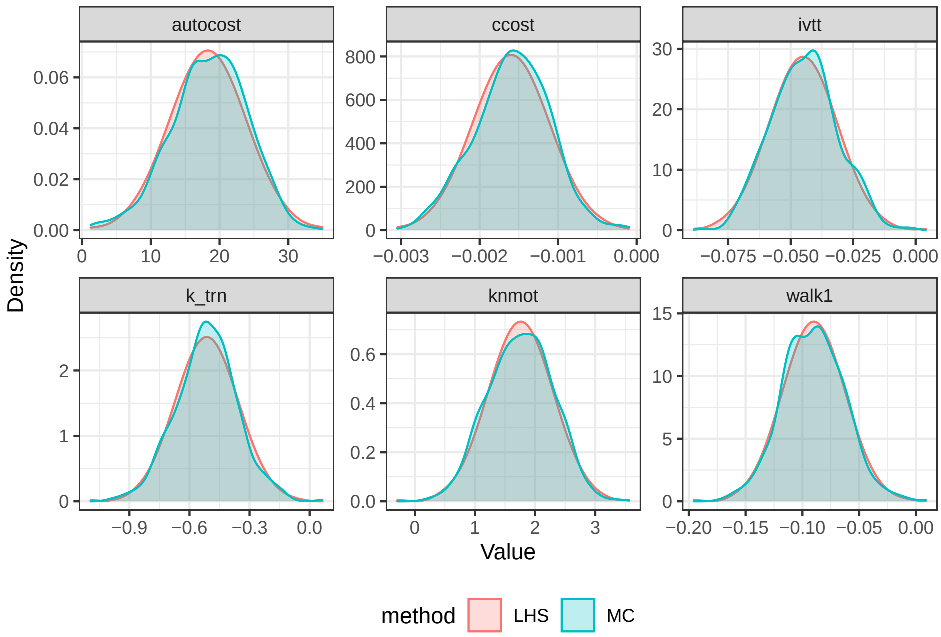

Figure 2: HBW Distributions for Input Parameters with 600 Draws

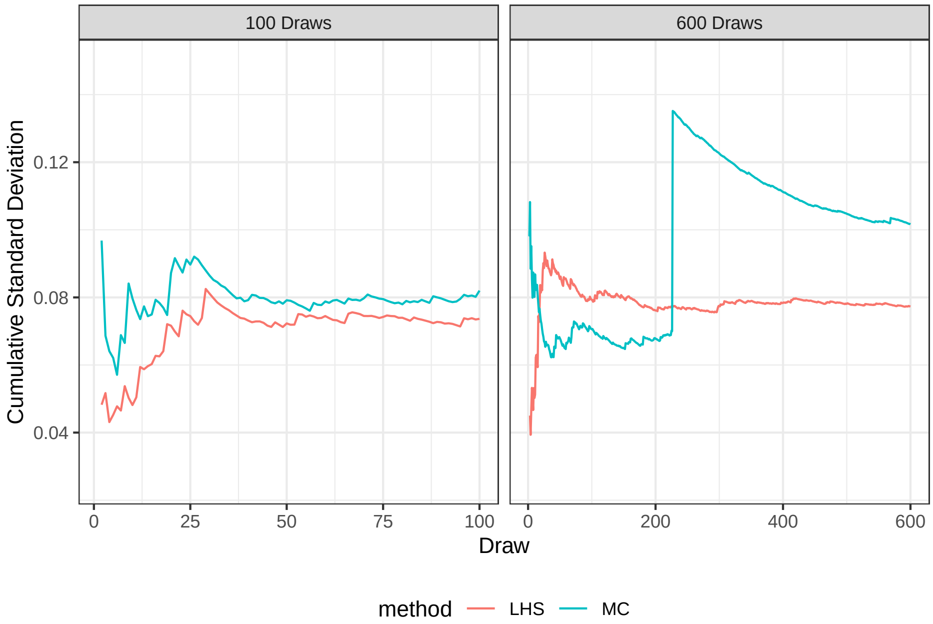

Figure 3: HBW Mean Logsum Standard Variation with 100 and 600 Draws

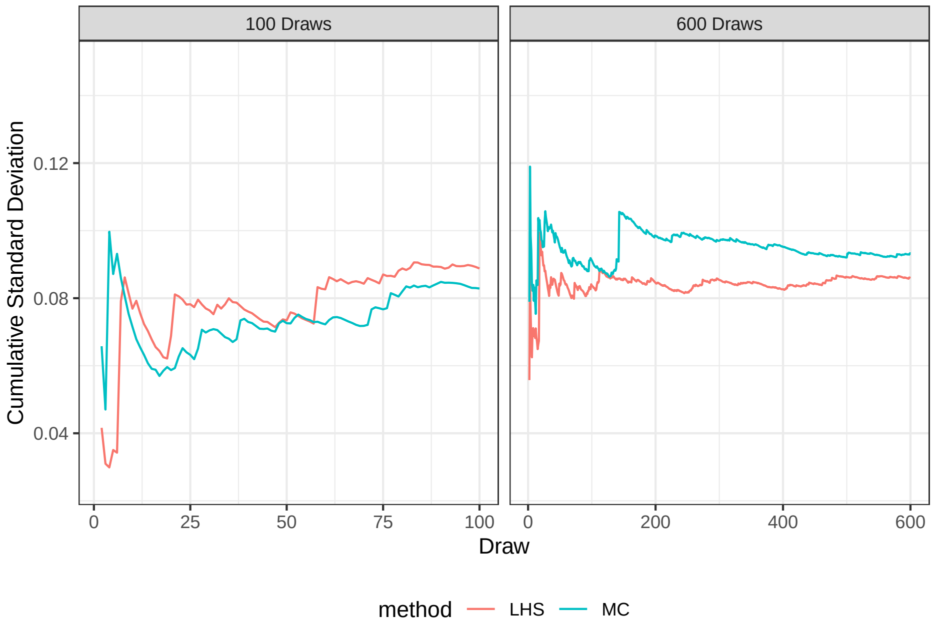

Figure 4: HBO Mean Logsum Standard Variation with 100 and 600 Draws

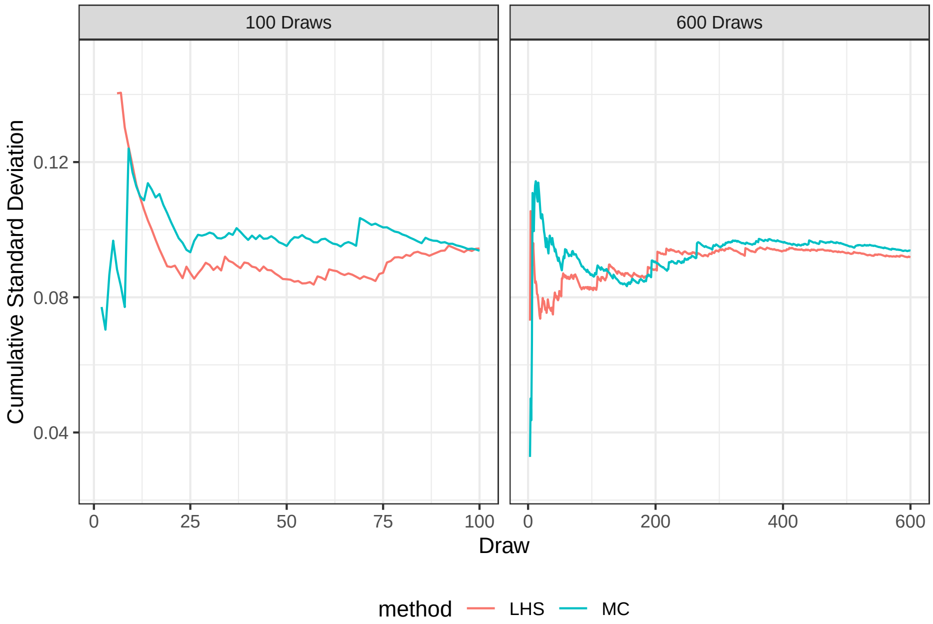

Figure 5: NHB Mean Logsum Standard Variation with 100 and 600 Draws

For all three trip purposes, the LHS method had its standard deviation stabilized between 100 and 200 draws. The MC method had still not stabilized to the same extent after 600 draws. This shows us that Latin Hypercube Sampling greatly decreases the iterations needed to approximate random sampling methods. Since LHS captures the possible variance at a small enough amount of iterations, it can and will be used for a statewide model.

References

Yang, C., Chen, A., Xu, X., and Wong, S. (2013). Sensitivity-based uncertainty analysis of a combined travel demand model. *Transportation Research Part B: Methodological*, 57:225–244.

Zhao, Y. and Kockelman, K. M. (2002). The propagation of uncertainty through travel demand models: an exploratory analysis. *The Annals of regional science*, 36(1):145–163.