Extended Abstract

Computation and Estimation of Path Travel Time Variability with Sparse Vehicle Trajectory Data

Monika Filipovska

Graduate Student Researcher

Email: monika.filipovska@u.northwestern.edu

Hani S. Mahmassani, Corresponding Author

William A. Patterson Distinguished Chair in Transportation

Director, Transportation Center

Email: masmah@northwestern.edu

Keywords: Routing; Stochastic networks; Least-time paths; Reliable paths; Travel time reliability

Word count: *1,196 (Excluding References) + 2 Figures*

INTRODUCTION

The performance of transportation networks is susceptible to fluctuations caused by the interactions between a range of factors, including the presence of work zones, traffic incidents, weather conditions, operation of control devices, variations in demand, etc. (Mahmassani et al., 2014). Such fluctuations cause uncertainty in the transportation network and as a result the travel time between two points in the network is better represented as a random variable with a non-stationary distribution. Broadly, networks characterized by such travel times are called stochastic, time-varying (STV) networks (Miller-Hooks and Mahmassani, 2000). Additionally, the structure of the network itself, the presence of travel patterns and the inherent dependence between traffic conditions on nearby links impose spatial and temporal dependencies between travel times in the network.

While estimates of road travel times are common and central to both performance evaluation and decision-making in transportation network, information about the reliability of travel times are rarely considered. Information on the variability of travel time can be used for risk-averse routing decisions or performance measurement or evaluation of the transportation networks. Therefore, knowledge of the variability (or conversely, the reliability) of travel times along routes in the network is more valuable than deterministic predictions of travel times for a host of applications.

This paper presents statistical and simulation-based approaches to computing the distribution of travel times on an arbitrary route in the network for a given departure time. In large-scale transportation networks, the number of possible routes in exponentially large, and an arbitrarily selected path may have few, if any, observations for full trip traversals in a given historical data set. Therefore, the approaches presented in this work are aimed at synthesizing sparse vehicle trajectory data at the link or sub-path level, while capturing the spatio-temporal dependencies of link travel times, in order to obtain accurate estimates for the travel time distribution along an arbitrary path in the network.

METHODOLOGY

The methodology presented in this paper consists of two components. Firstly, simulation-based approaches for the estimation of the convolution of link or segment travel time distributions with spatio-temporal dependencies are presented. Subsequently, a combinatorial data-mining search approach for synthesizing trajectory data is proposed. The former simulation-based estimation approaches are general and can be utilized for the convolution of link or segment-level travel time distributions, including the path segmentations resulting from latter approach. The segment search approach is intended be used as an alternative to methods that involve correlation estimation, since it aims to implicitly capture spatial correlations so that simple convolution estimation methods, i.e. assuming spatial independence, would be sufficient.

The Monte-Carlo simulation-based approaches for convolution estimation presented in this paper are as follows:

- Assuming independent random segment travel times
- Assuming random segment travel times with time-varying distributions

- Assuming random correlated link travel times with time-varying distributions
- Assuming random correlated link travel times with time-varying distributions and covariances

The combinatorial data-mining segment search approach is a heuristic method for determining groups of non-overlapping segments that fully cover the desired path that can be used to estimate the path travel time distribution. The procedure, presented in Figure 1, is applied iteratively in order to make use of any potentially remaining data. The resulting segmentations of the path can then be used as the basis of any of the Monte-Carlo simulation-based approaches presented above. However, the additional step of segment search is intended to eliminate the burden of estimating and accounting for correlation and as such the results of this approach are intended to be applied with the first two of the Monte-Carlo approaches.

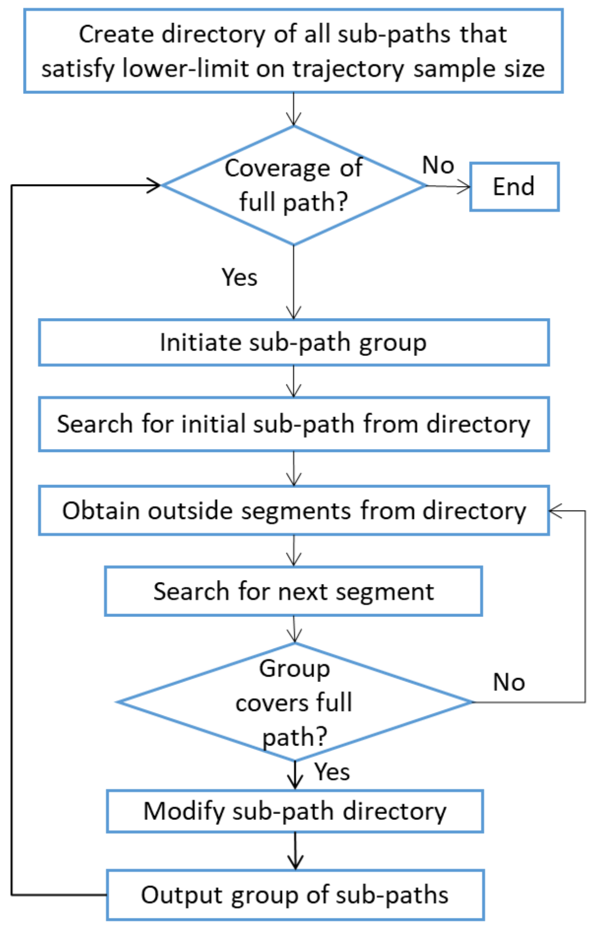

Figure 1. Diagram illustrating the combinatorial segment search approach.

The approaches presented above are compared to a few methods from the literature on probability modeling and estimation that can be used for the estimation of path travel time distributions with correlation. The first is based on a Markov chain Monte Carlo sampling method to simulate multivariate distributions known as the Metropolis-Hastings algorithm. Detailed exposition of the Metropolis-Hastings algorithm itself can be found in (Chib and Greenberg, 1995). The second approach is a commonly used approximation for the sum of correlated non-negative random variables via a lognormal distribution (Mehta et al., 2007).

RESULTS

The presented methods for the estimation of path travel time distributions on an urban network were tested using simulated trajectory data on the network of Chicago, with 1578 nodes and 4805 links. A total of 25 analysis scenarios were simulated based on daily demand and weather conditions in order to provide variation in travel times in the STV network. The simulated data set contains full trajectory information which was used for estimating travel time distributions. The detailed calibration and other characteristics of this data set are out of the scope of this study and can be found in (Yelchuru et al., 2017). For this study, the trajectories observed in the morning peak period from 6 AM to 9 AM were considered. Additionally, based on operational scenarios defined in the simulation and observed variations in travel times in the resulting data, travel time distributions were estimated conditional on 4 weather conditions: clear weather, rain, snow, and low visibility. In order to test the estimation accuracy of the presented methods, a set of 35 paths were selected from the network. The paths were selected so as to have a large number of vehicle traversals, i.e. trajectory observations, across all four weather conditions, so that they could provide a reliable estimate of the path travel time distribution to be used as the ground truth.

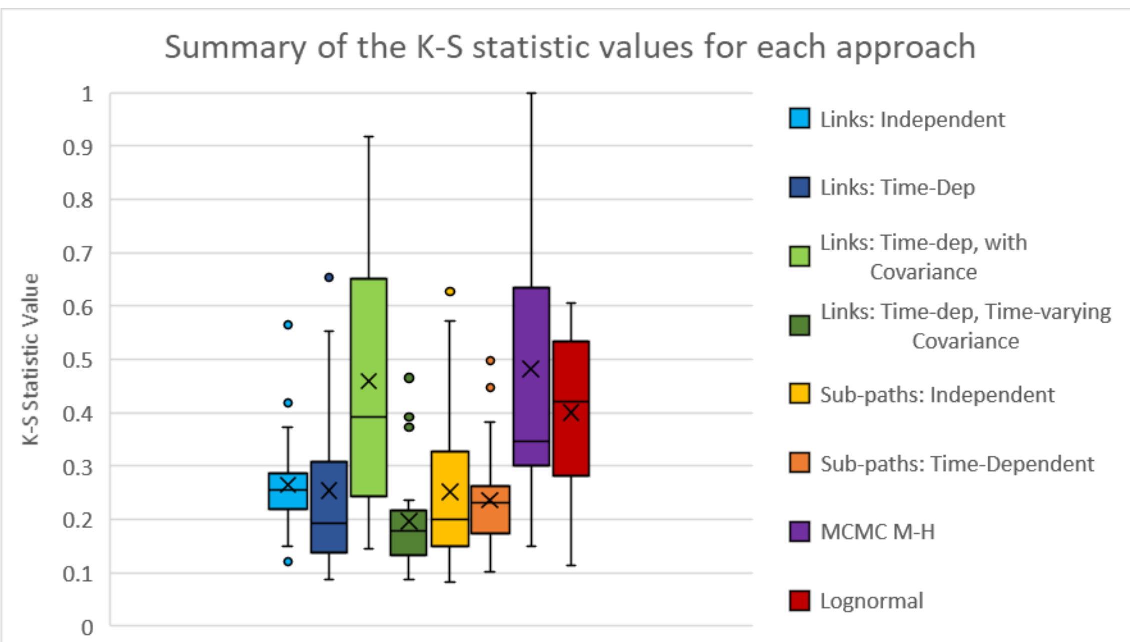

Figure 2. Box Plot of the K-S statistic values for all of the estimation approaches

The summary results in Figure 1 show the variation of K-S statistic values for each of the methods. The approaches yielding the least variation in the K-S statistic value were the link-based approach assuming independence, the link-based approach assuming time-dependence and with a time-varying covariance structure, and the sub-paths approach with time-dependence. The approach with the most variation in the K-S statistic value was the links-based approach with time-dependence and a stationary covariance structure. This approach results in some of the highest values of the K-S statistic, but also in some cases outperforms some of the other approaches. Since our data showed the time-varying property of the covariance between link travel-times, assuming stationary covariance can lead to a biased distribution in a lot of cases.

CONCLUSION

This paper addresses the question of estimating path travel time distributions in stochastic time-varying networks with generalized correlation structures. The problem is addressed in two manners. Firstly, building on existing Monte Carlo simulation approaches, MCS approaches for random variables with time-varying distributions are presented for the cases assuming no correlations, stationary correlations, and time-varying correlations between link travel times. Secondly, to bypass the estimation of generalized time-varying correlations in an STV network, an alternative data-mining approach is presented for synthesizing trajectory data at the segment level.

ACKNOWLEDGEMENTS

This paper is based on research motivated in part by a research project funded by the Federal Highway Administration (FHWA), US Department of Transportation. The authors have benefited from insight provided by Douglas Laird at FHWA, David Hale at Leidos, Inc., Monty Abbas at Virginia Tech, and Archak Mittal at Ford Research. The authors remain responsible for all content of the paper.

REFERENCES

Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. Am. Stat. 49, 327–335.

Mahmassani, H.S., Kim, J., Chen, Y., Stogios, Y., Brijmohan, A., Vovsha, P., 2014. Incorporating reliability performance measures into operations and planning modeling tools. Transportation Research Board.

Mehta, N.B., Wu, J., Molisch, A.F., Zhang, J., 2007. Approximating a sum of random variables with a lognormal. IEEE Trans. Wirel. Commun. https://doi.org/10.1109/TWC.2007.051000

Miller-Hooks, E., Mahmassani, H.S., 2000. Least Expected Time Paths in Stochastic, Time-Varying Transportation Networks. Transp. Sci. 34, 198–215. https://doi.org/10.1287/trsc.34.2.198.12304

Yelchuru, B., Raj Kamalanathsharma, Mahmassani, H.S., Hong, Z., Mittal, A., Xu, X., 2017. Analysis, modeling, and simulation (ams) testbed framework for dynamic mobility applications (dma) and active transportation and demand management (atdm) programs. US Department of Transportation.