Transportation Research Part B 000 (2017) 1–20

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Stochastic travel demand estimation: Improving network identifiability using multi-day observation sets

Yudi Yang[a], Yueyue Fan[a,*], Roger J.B. Wets[b]

[a] Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, United States
[b] Department of Mathematics, University of California, Davis, CA 95616, United States

A R T I C L E I N F O

Article history:
Received 15 April 2017
Revised 14 October 2017
Accepted 16 October 2017
Available online xxx

Keywords:
Stochastic O-D estimation
Traffic network
Domain knowledge
Identifiability

A B S T R A C T

Stochastic travel demand estimation is essential to support many resilience and reliability based transportation network analyses. The problem of estimating travel demand based on sensor data often results in an ill-posed inverse problem, where solution uniqueness cannot be ensured. To overcome this challenge, effective utilization of more information/data, preferably from reliable sources, becomes critical. Conventional demand estimation methods often sacrifice system structural information during the process of compressing sensor data into its statistics. Loss of structural information, which captures critical relation between observed and estimated parameters, inevitably causes more dependence on unrealistic assumptions and unreliable data. Our model is designed to preserve all structural information contained from different observation sets and allow it to directly contribute to the identification of population parameters of travel demand. The proposed hierarchical framework integrates two traditionally distinctive identification problems, mean demand estimation and trip table reconstruction. Through mathematical analyses and numerical experiments, we show that the proposed framework improves parameter identifiability and leads to better estimation quality compared to conventional methods. The proposed framework is also flexible to accommodate a wide variety of travel behavior assumptions and estimation principles. As an example among many possible alternatives, Wardrop equilibrium based traffic assignment and generalized least square are implemented and tested using a case study based on a moderately large network.

1. Introduction

Origin-Destination (O-D) travel demand is critical information needed for planning and operating a transportation system. In the past decades, most O-D demand estimation studies focused on producing a single deterministic estimation of O-D matrix. On the other hand, as reliability and resilience become a significant concern in many applications, a large number of transportation system studies now call for a more comprehensive understanding of demand stochasticity. For example, extensive efforts were made to model travel time variability under day-to-day variation (e.g. Clark and Watling, 2005), to analyze traffic equilibrium considering late-arrival risk (e.g. Zhou and Chen, 2008), to design dynamic vehicle routing scheme based on reliability (e.g. Fan et al., 2005), and to develop resilient network land use strategy (e.g. Yim et al., 2011). The list

* Corresponding author.
E-mail address: yyfan@ucdavis.edu (Y. Fan).

https://doi.org/10.1016/j.trb.2017.10.007
0191-2615/© 2017 Elsevier Ltd. All rights reserved.

of publications that assume the availability of stochastic demand information could go much longer. To the contrary, the literature on estimating stochastic demand is by far limited.

Recent attempts to estimate stochastic travel demand usually concern the same observational period of a few hours for multiple days. Sensor data collected within the period is aggregated to a single vector of observed flows, which is called an *observation set*. A typical estimation horizon could span over a year. Lo et al. (1996) first introduced the concept of population parameter in modeling stochastic travel demand. Vardi (1996) derived a full likelihood function of parameters given repeated measurements considering the correlation between link counts. Substantial efforts were made in improving statistical approaches designed for stochastic O-D estimation, including differentiating reconstruction, mean estimation and prediction problems (Hazelton, 2001), developing Bayesian approaches for inferring time-varying O-D matrices (Hazelton, 2008a), and considering an integer valued stochastic demand inference problem (Hazelton, 2015). These studies took into account a simple linear mapping between observed traffic flows and variables of interest, O-D or path flows, based on physical network topology; domain knowledge reflecting travelers' route choices over a congested network was not incorporated in those estimation problems.

Most studies on O-D estimation that utilized domain knowledge regarding travelers' network routing behaviors aimed at pinpointing a vector of demand that best matches a single observation set. For example, Yang et al. (1992) and Maher et al. (2001) reconstructed an O-D matrix that could produce the closest link flows according to user equilibrium (UE) and stochastic user equilibrium (SUE) respectively. An exception was seen in a recent study by Shao et al. (2014), who formulated stochastic demand estimation with equilibrium as a bi-level program, with the upper level featuring a quadratic problem of the difference between sample statistics and estimated statistics and the lower level imposing reliability based user equilibrium (RUE)[1]. In an online travel demand estimation context, dynamic traffic assignment model is typically used to capture the relationship between observed and estimation parameters, and Kalman Filter or its variations are exploited to provide sequential update of the dynamic travel demand (Antoniou et al., 2006; Carrese et al., 2017).

A major challenge encountered in almost all O-D estimation problems is observability/identifiability issue. In the context of estimating stochastic demand, the term observability refers to the determinacy of short-term O-D flows from a single observation set, while identifiability refers to the uniqueness of parameters that govern underlying probability distribution of travel demand. In brief, observability is associated with a reconstruction problem and identifiability is with an estimation problem. Several studies have investigated observability issues in O-D matrix reconstruction problems (Castillo et al., 2008a and Viti et al., 2014), which assumed linear relationship between O-D flows and observed values. Castillo et al. (2015) provided a thorough review of flow observability and O-D demand estimation problems, and suggested an integrative view for approaching these seemingly different problems using a unified mathematical optimization framework. In the area of dynamic O-D demand estimation, Balakrishna (2006) discussed possible variations in observability with different types of data, including counts, speed and density.

When it comes to coping with the identifiability issue for stochastic travel demand estimation, researchers typically relied on additional information provided by specific statistical or assignment models. For example, Vardi (1996) took advantage of the unique feature of a Poisson distribution, that is, mean being equal to variance, and Hazelton (2003) assumed variance as a monomial function of mean to improve the identifiability. Singhal and Michailidis (2007) provided sufficient conditions of n-order moments under a latent variable model that captures the flow correlation in a data network that has a similar structure. Shao et al. (2014) assumed multivariate normal distribution for demand and made good use of RUE mapping since its input-output relation is expressed directly using compressed statistics of mean and variance. There were pioneer studies (Cremer and Keller, 1987; Bell, 1991) investigating the potential of improving identifiability via exploiting structural information from different time-of-day observations in the context of deterministic dynamic demand estimation. In fact, in stochastic demand estimation, the general philosophy of utilizing potentially dissimilar data-dependent information from multiple sets is also promising and should be applicable under a large variety of model assumptions, albeit rarely seen in the literature to the best of our knowledge.

The aim of this paper is to provide a general strategy attending to the identifiability issue of stochastic travel demand estimation, given multi-day observation sets where the O-D flow is not fully observable in some/all of the sets. We propose a stochastic demand estimation framework that can fully utilize the information from link and path data as well as network topology and traffic operation knowledge. Instead of being restricted by a special probabilistic model or an assignment rule, it is versatile to accommodate a large variety of modeling assumptions. In addition to estimating the population parameters of travel demand probability distribution, the O-D matrix associated with each observation set is simultaneously reconstructed in this framework.

The rest of the paper is organized as follows. The second section introduces network models and data types in a stochastic travel demand estimation context. The third section presents the new modeling framework in comparison with conventional methods and provides detailed observability and identifiability analyses. The fourth section demonstrates how the proposed framework can be implemented to solve a stochastic estimation problem with a specific estimation criterion and assignment rule. The fifth section provides a numerical study on a moderately large network. The last section concludes the paper with discussion and possible future extensions.

[1] Given the mean vector and covariance matrix of O-D demand, RUE assignment model generates the mean and covariance of link flows.