Transportation Research Part C 88 (2018) 227–256

Contents lists available at ScienceDirect

Transportation Research Part C

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

Statistical inference of probabilistic origin-destination demand using day-to-day traffic data

Wei Ma[a], Zhen (Sean) Qian[a,b,*]

[a] Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[b] H. John Heinz III Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, United States

ARTICLE INFO

Keywords:
Probabilistic origin-destination demand estimation
Statistical traffic assignment
Probability distribution
Variance decomposition
Data driven

ABSTRACT

Recent transportation network studies on uncertainty and reliability call for modeling the probabilistic O-D demand and probabilistic network flow. Making the best use of day-to-day traffic data collected over many years, this paper develops a novel theoretical framework for estimating the mean and variance/covariance matrix of O-D demand considering the day-to-day variation induced by travelers' independent route choices. It also estimates the probability distributions of link/path flow and their travel cost where the variance stems from three sources, O-D demand, route choice and unknown errors. The framework estimates O-D demand mean and variance/covariance matrix iteratively, also known as iterative generalized least squares (IGLS) in statistics. Lasso regularization is employed to obtain sparse covariance matrix for better interpretation and computational efficiency. Though the probabilistic O-D estimation (ODE) works with a much larger solution space than the deterministic ODE, we show that its estimator for O-D demand mean is no worse than the best possible estimator by an error that reduces with the increase in sample size. The probabilistic ODE is examined on two small networks and two real-world large-scale networks. The solution converges quickly under the IGLS framework. In all those experiments, the results of the probabilistic ODE are compelling, satisfactory and computationally plausible. Lasso regularization on the covariance matrix estimation leans to underestimate most of variance/covariance entries. A proper Lasso penalty ensures a good trade-off between bias and variance of the estimation.

1. Introduction

Origin-destination (O-D) demand is a critical input to system modeling in transportation planning, operation and management. For decades, O-D demand is deterministically modeled, along with deterministic models of link/path flow and travel cost/time in classical traffic assignment problems. Transportation network uncertainty and reliability call for modeling the stochasticity of O-D demand, namely its spatio-temporal correlation and variation. With the increasing quantity and quality of traffic data collected years along, it is possible to learn the stochasticity of O-D demand for better understanding stochastic travel behavior and stochastic system performance metrics. Some studies considered the stochastic features of O-D demand, but few estimated the mean and variance of O-D demand from day-to-day traffic data. What is missing in the literature is the capacity to estimate spatially correlated multivariate O-D demand, in conjunction with a sound network flow theory on probabilistic route choices that can be learned from day-to-day traffic data. In view of this, this paper develops a novel data-friendly framework for estimating the mean and variance/covariance of O-D demand based on a generalized statistical network equilibrium. The statistical properties towards the estimated probabilistic O-D

* Corresponding author at: Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States.
E-mail addresses: weima@cmu.edu (W. Ma), seanqian@cmu.edu (Z.S. Qian).

https://doi.org/10.1016/j.trc.2017.12.015
Received 24 August 2017; Received in revised form 5 December 2017; Accepted 24 December 2017

demand are also analyzed and provided. The process of O-D demand estimation (ODE) is further examined in real-world networks for insights.

O-D estimation (ODE) requires an underlying behavioral model, based on which O-D demand is estimated such that it best fits observations. Behavioral models in the static network context, namely route choice models, are also known as static traffic assignment models. The classical traffic assignment models (e.g., Dafermos and Sparrow, 1969; Patriksson, 1994) deterministically map the deterministic O-D demand $q \in \mathbb{R}_+^{R\times S}$ to link flow $x \in \mathbb{R}_+^N$ or path flow $f \in \mathbb{R}_+^K$ (where R,S,N and K are the cardinality of sets of origins, destinations, links, and paths, respectively). In fact, the deterministic O-D demand q is assumed to represent the *mean* number of vehicles in the same peak hour from day to day. Likewise, link (path) flow is also deterministic, representing the *mean* number of vehicles on a link (path) in the same hour from day to day. The classical traffic assignment models (such as User Equilibrium and Stochastic User Equilibrium) lay out the foundation of deterministic O-D estimation methods, namely to estimate q in a way to best fit observed data related to a subset of link/path flow x,f. Deterministic O-D estimation (ODE) include the entropy maximizing models (Van Zuylen and Willumsen, 1980), maximum likelihood models (Spiess, 1987; Watling, 1994), generalized least squares (GLS) models (Cascetta, 1984; Bell, 1991; Yang et al., 1992; Wang et al., 2016), Bayesian inference models (Maher, 1983; Tebaldi and West, 1998) and some recent emerging combined models (Ashok and Ben-Akiva, 2002; Castillo et al., 2008c, 2014a; Li, 2009; Lu et al., 2015; Woo et al., 2016). For more details, readers are referred to the comprehensive reviews by Bera and Rao (2011) and Castillo et al. (2015).

Classical traffic assignment models and ODE methods overlook the variance/covariance of demand and link/path flow, an essential feature for network traffic flow. Recent studies on network reliability and uncertainty model stochasticity of the network flow. Since ODE requires a traffic assignment model, we first review statistical traffic assignment models, followed by ODE models that take into account stochasticity.

One aspect of the stochastic network flow is using probability distributions to represent O-D demand. For example, Waller et al. (2006) and Duthie et al. (2011) sampled O-D demand from given multivariate normal distributions (MVN) and evaluated the network performance under classical User Equilibrium (UE) condition. Chen et al. (2002) used a similar simulation-based method to evaluate travelers' risk-taking behavior due to probabilistic O-D demand. All these studies indicate the O-D variation is of great importance to network modeling and behavioral analysis. Statistical traffic assignment models consider various O-D probability distributions such as Poisson distributions (Clark and Watling, 2005), MVN (Castillo et al., 2008b), multinomial distributions (Nakayama, 2016). Nakayama and Watling (2014) summarized different formulations and proposed a unified framework for stochastic modeling of traffic flows. Advantages and disadvantages of modeling traffic with those probabilistic distributions are also discussed by Castillo et al. (2014b). Shao et al. (2006a,b) proposed a reliability-based traffic assignment model (RUE) and extended it to consider different travelers' risk taking behavior. Lam et al. (2008) further extended the model to consider the traffic uncertainty and proposed reliability-based statistical traffic equilibrium. Zhou and Chen (2008) and Chen and Zhou (2010) proposed a α-reliable mean-excess traffic assignment model which explicitly models the travel time distribution and consider the reliability and uncertainty of the travel time on travelers' route choice behavior. Other studies (Haas, 1999; Waller et al., 2001; Duthie et al., 2011) show that the variance/covariance matrix of the O-D demand have a significant influence on network traffic conditions.

Though adopting stochastic O-D demand, those traffic assignment models (except Clark and Watling (2005)) assumed non-atomic (infinitesimal) players and therefore are unable to capture the stochasticity of route choices that vary from day to day (the proof is shown in Ma and Qian (2017)). Classical UE, SUE and RUE are all deterministic route choice models where the number of (infinitesimal) players assigned to each route is fixed, rather than being stochastic. Thus, those models are unable to explain the day-to-day variation of observed traffic counts at the same location and the same time of day. To further see how classical equilibrium models overlook the route choice stochasticity, suppose there is Q travelers where Q is a random variable to be realized on each day. Given the probability of choosing a route p, the route flow is deterministically identified by the number of infinitesimal users who take this route, $F = pQ$. Even if the route choice probability p is determined by stochastic choice models (such as Probit, Logit, etc.), these models still assume that a fixed number Qp of travelers take this route on each day, as a result of non-atomic equilibrium. This does not, theoretically, allow to model the day-to-day variations of travelers' choices. Recent studies on statistical traffic assignment models indicate that the route flow is the aggregation of random choices of O-D demand, and thus also random (Watling, 2002a,b; Nakayama and ichi Takayama, 2006; Nakayama and Takayama, 2003; Nakayama and Watling, 2014). Travelers' route choice follows a multinomial distribution with the probability obtained from route choice models, $f \sim \text{Multinomial}(Q,p)$. To distinguish to what extent stochasticity is modeled for route choices, we refer to the former (classical) models as "*fixed portions with stochastic route choice models*" and the latter models as "*probabilistic distributions with stochastic route choice models*". Though route choices are stochastic, these studies did not work directly with the covariance of demand among all O-D pairs. A detailed comparison of those assignment models is further illustrated in Ma and Qian (2017).

Given a route choice probability, we can derive probability distributions of path/link flow. However, it is non-trivial to establish a statistical network equilibrium where the route choice probability is determined endogenously as a result of stochastic O-D demand, path/link flow and network conditions. In the deterministic settings, UE, SUE or RUE simultaneously determines the *mean* path/link flow, and the route choice probability (e.g., Yang et al., 1992). Very few studies examined the statistical network equilibrium. Davis and Nihan (1993) proposed a Markov process to model the day-to-day variation of traffic flow given a fixed regional population. The stochastic route choice on a particular day is assumed to be related to the stochastic network condition of the previous days (Cascetta and Cantarella, 1991). When the network evolves from day to day, there exists a network equilibrium where stabilized probability distributions of path/link flow are reached. Different from this approach, Ma and Qian (2017) proposed a generalized statistical equilibrium where each traveler makes stochastic choices based on his/her entire past experience, namely the probability distributions of the equilibrated network conditions. Ma and Qian (2017) integrates multivariate probability distributions of O-D