Link Travel Time Inference Using Entry/Exit Information of Trips on a Network

Kai Yin*, Wen Wang †, X. Bruce Wang †, and Teresa M. Adams ‡

July 6, 2015

Abstract

This paper studies link travel time estimation using entry/exit time stamps of trips on a steady-state transportation network. We propose two inference methods based on the likelihood principle, assuming each link associates with a random travel time. The first method considers independent and Gaussian distributed link travel times, using the additive property that trip time has a closed-form distribution as the summation of link travel times. We particularly analyze the mean estimates when the variances of trip time estimates are known with a high degree of precision and examine the uniqueness of solutions. Two cases are discussed in detail: one with known paths of all trips and the other with unknown paths of some trips. We apply the Gaussian mixture model and the Expectation-Maximization (EM) algorithm to deal with the latter. The second method splits trip time proportionally among links traversed to deal with more general link travel time distributions such as log-normal. This approach builds upon an expected log-likelihood function which naturally leads to an iterative procedure analogous to the EM algorithm for solutions. Simulation tests on a simple nine-link network and on the Sioux Falls network respectively indicate that the two methods both perform well. The second method (i.e., trip splitting approximation) generally runs faster but with larger errors of estimated standard deviations of link travel times.

---

*Nomis Solutions, Inc. `yinkai1000@gmail.com`

†Zachry Department of Civil Engineering, Texas A&M University, `{ww051213,bwang}@tamu.edu`. Bruce Wang is the corresponding author.

‡Department of Civil and Environmental Engineering, University of Wisconsin - Madison, `adams@engr.wisc.edu`

1 INTRODUCTION

Travel time is one of the most important factors when a traveler plans a route from an origin to a destination, and is also critical to transportation planners and operators as a performance measure. Accurate travel time estimation on a transportation network is therefore becoming an essential task and is made possible now by widely available traffic data.

Travel time data on a network is regularly obtained from traffic tracking such as through probing phones (Bar-Gera, 2007; Ygnace et al., 2000), global positioning system (GPS) devices (Bertini and Tantiyanugulchai, 2004), and vehicle ID readers (through Bluetooth or vehicle plate identification and matching, e.g., Haghani et al. (2010); Chang et al. (2004)). While the associated methods to estimate roadway travel time range from regression models (Chan et al., 2009), machine learning approaches (Zheng and Van Zuylen, 2013), and to analytical models dealing with traffic conditions (Hellinga et al., 2008), many required parameters limit their applicability in practice and there is a lack of general model approach to the network-wide travel time estimation problem.

Valid statistical analysis becomes increasingly important as the data becomes widely available (Fan et al., 2014). In what follows, we will briefly review two categories of statistical approaches in relevant literature: the traditional maximum-likelihood method and the Bayesian approach.

Among the scant literature that focuses directly on this topic, Hunter et al. (2009) formulate a maximum-likelihood problem to estimate link travel time distributions on an arterial network. Their model considers the observations of unknown trajectories. They present an Expectation-Maximization (EM) algorithm to simultaneously learn the likely paths of probe vehicles as well as the travel time distributions on the network. They assume that travel times on different links are independent and briefly conduct numerical tests using San Francisco taxi data. Instead of the assumption of independent link travel times, Jenelius and Koutsopoulos (2013) present a statistical model for travel time estimation on an urban road network considering the correlation between travel times on different links. They capture the correlation using a moving average specification for link travel times. The specific information of link attributes (such as speed limit and roadway functional class) and trip conditions (such as day-of-week, time-of-day, and weather condition) are incorporated as explanatory variables. The model is estimated using maximum-likelihood method, and is applied to a particular route on the Stockholm network in Sweden.

In contrast to the traditional maximum-likelihood method, some studies apply the Bayesian approach to travel time distribution prediction. Hofleitner et al. (2012a,b) propose a dy-

namic Bayesian network for unobserved traffic conditions and model link travel time distributions conditional on traffic state. Their method is from the conventional traffic flow perspectives, and is applied to a San Francisco road network to predict travel times using taxi data. Westgate et al. (2013) also propose a Bayesian model to estimate the distribution of ambulance travel times on road segments in Toronto. They apply a multinomial Logit model to formulate the path choices for ambulance trips, and perform the path inference and travel time estimation simultaneously using a Bayesian approach. They also assume that link travel times are independent and log-normally distributed. The parameters are estimated using Markov Chain Monte Carlo (MCMC) methods. Instead of modeling travel times at the link level in the previous work, Westgate et al. (2014) model ambulance travel times at the trip level. They propose a regression approach for estimating the ambulance travel time distribution on an arbitrary route, and use a Bayesian formulation to estimate the model parameters. The advantage of applying the Bayesian approach is that it utilizes expert knowledge as prior information, and tackles many complicated problems that traditional statistical approaches find difficult to analyze. However, its implementation relies on computationally expensive methods such as the MCMC.

This research aims to develop inference methods for link travel time estimation on a steady state network, assuming that each link is associated with a random travel time due to different traveling vehicles and variation of day time. We believe the statistical characteristics of the link travel time as well as the distribution can be measured approximately. We estimate network-wide link travel times by only using vehicle start and end locations and time of trips, referred to as traveler entry/exit time stamps in this paper. This type of data is representative of the data available when discrete points of a trip are recorded. Sparse vehicle trajectories reported by GPS-equipped probe vehicles or smart phones (Wang et al., 2014) can also be regarded as a particular case of traveler entry/exit trip information on a network. By trip we mean sequential records that denote the time stamps for start and end nodes on a roadway network. For example, two consecutive GPS records are considered as a trip in this study, though they essentially represent a trip segment. Specially, this research is motivated by a practical application on a toll road network, in which traveler entry/exit time stamps are recorded at tollbooths and the toll road authority has a practical need to use the travel time inference results to evaluate the toll systems. Other potential applications include using public transit data for network performance analysis when passenger entry/exit information is recorded at fare boxes (Ma et al., 2013).

We start with the assumption of independent and normally distributed link travel times, and present the EM algorithm to address the trips with unknown paths, as in Hunter

et al. (2009) and Siripirote et al. (2013). However, our study differs from this earlier work by focusing on exploring the analytical properties of the proposed methods. We examine the impact of errors in trip variance estimates on mean link travel time estimates, and investigate the uniqueness of solutions in the algorithm. We also provide confidence intervals for mean link times.

Furthermore, we propose a statistical method of trip splitting approximation, as Method II in this paper, to mainly address a technical situation in which the summation of random link travel times for a path does not have a closed-form probability distribution. We assume that trips on the same path under similar traffic conditions (e.g., different vehicles are measured within a time interval of 30 minutes or so) have approximately a constant proportion of the trip time for a traversed link. Similar idea of decomposing trip travel time has already been proposed in practical applications (Hellinga et al., 2008), but without appropriate justification and investigation. The proposed trip splitting approximation method may be applied to arbitrary distributions, and is statistically justified for the network estimation problem. Our numerical tests indicate that this method is computationally cheaper, though it may lead to relatively larger errors for link time standard deviation estimates. Its potential application would be more promising if more traffic information is available.

The remainder of this paper is organized as follows. Section 2 proposes the first method assuming that the trip time has a closed-form distribution, using Gaussian distribution for link travel time as an example. Two cases are discussed in detail: one with known paths of all trips and the other with unknown paths of some trips. Section 3 develops a statistical framework of the trip splitting method. Section 4 tests the proposed methods with simulated data on a simple nine-link test network and the Sioux Falls network respectively. Section 5 discusses the advantages and disadvantages of both methods. Section 6 concludes the paper.

2 METHOD I: ESTIMATION USING TRIP TIME DISTRIBUTIONS

In this study, a roadway network is represented by a graph where links (edges) stand for roadway segments and nodes represent connections of links (e.g., intersections or junctions). Each link is associated with a random travel time that follows a certain type of distribution. A path is defined as an alternating sequence of links and nodes from an origin to a destination node (known as an OD pair). Each trip consists of a path, the entry (starting) time at the origin, and the exit (ending) time at the destination. Multiple trips may take place on the same path. Each trip on the network has an OD pair with entry/exit times. Paths may not be known for some trips. A trip time, derived from the

difference between entry/exit times, is the summation of link travel times along a path. With a sufficiently large number of trips observed, our goal is to estimate the parameters of link travel time distribution by handling the unobserved paths if necessary.

In what follows, we develop models to address the case with known paths and the case with both known and unknown paths, respectively. Modeling the randomness of trip times in terms of the specific link time distributions presents a key technical challenge. In Method I that follows, we develop models for Gaussian distributed link travel times using the additive property that the summation of Gaussian random variables still follows a Gaussian distribution. Note that we generally assume all link travel times are independent in our study here unless specified otherwise.

2.1 Link Time Estimation Using Trips with Known Paths

We first study the basic case in which all the observed trips have known paths. In other words, each OD observation has a specific set of links on which the trip takes place. Link travel times are estimated according to a specific time interval of the day, although the time interval may be wide such as half an hour or longer.

We let A be the set of road links and n be the total number of links. Let I be the total number of observations and x_i denote the observed travel time of trip i. We assume that I is larger than n throughout the rest of the paper. The set of observations is represented by D, i.e., $D = \{x_1, x_2, \ldots, x_I\}$. As all the trips have known paths, we denote by $\delta_{i,a}$ an incidence indicator, which is equal to 1 when link a is on trip i and 0 otherwise. Let the corresponding incidence matrix be $\Delta = [\delta_{i,a}]_{I \times n}$. In addition, we denote the mean travel time on link a by μ_a and the corresponding standard deviation by σ_a, where $a \in A$. Let μ and σ be the n-by-1 vectors of μ_a and σ_a, respectively. Following Rakha et al. (2006) and Wen et al. (2014), we make the following assumption:

Assumption: All link travel times on the study network are independent and normally distributed random variables, as denoted by $\mathcal{N}(\mu_a, \sigma_a)$ for each link a.

Here, we maximize the following likelihood for the trip observations:

$$\text{Maximize} \quad \mathcal{LL}(\eta, \tau \mid D) \;\; = \;\; \sum_i \log\left(\frac{1}{\sqrt{2\pi}\tau_i} e^{-\frac{(x_i - \eta_i)^2}{2\tau_i^2}}\right), \tag{1}$$

where η_i and τ_i are respectively the mean travel time and the standard deviation of trip

i. η and τ are vectors of η_i and τ_i, respectively. The following equations hold:

$$\eta_i = \sum_a \delta_{i,a}\mu_a, \tag{2}$$

$$\tau_i = \sqrt{\sum_a \delta_{i,a}\sigma_a^2}. \tag{3}$$

The objective function (1) is equivalent to a minimization function as follows

$$\text{Minimize} \quad \mathcal{W}(\mu, \sigma \mid D) = \sum_i \left(\log\left(\sum_a \delta_{i,a}\sigma_a^2\right) + \frac{1}{\sum_a \delta_{i,a}\sigma_a^2}\left(x_i - \sum_a \delta_{i,a}\mu_a\right)^2 \right). \tag{4}$$

The objective leads to the following equations, by setting the partial derivative to zero for a specific link a with respect to its parameters μ_a and σ_a, respectively,

$$\sum_i \frac{\delta_{i,a}(\sum_b \delta_{i,b}\mu_b)}{\sum_b \delta_{i,b}\sigma_b^2} = \sum_i \frac{\delta_{i,a}x_i}{\sum_b \delta_{i,b}\sigma_b^2}, \text{ for any link } a, \tag{5}$$

$$\sum_i \frac{\delta_{i,a}}{\sum_b \delta_{i,b}\sigma_b^2} = \sum_i \left(\frac{\delta_{i,a}}{(\sum_b \delta_{i,b}\sigma_b^2)^2}\left(x_i - \sum_b \delta_{i,b}\mu_b\right)^2 \right), \text{ for any link } a. \tag{6}$$

Equations (5) and (6) are nonlinear but one may resort to Newton–Raphson's method for solutions.

2.1.1 Matrix representation

It is convenient to format Equations (5) and (6) in matrix form to simplify the further analysis. Two approaches are available: one through the observation–link incidence matrix and the other through the path–link matrix. While the former appears more natural, the latter has a more compact form that will be useful for practical implementation. We present the first approach in this section and present the second in Appendix A.

Let X be a I-by-1 vector with the i-th element being x_i and Σ be the $n \times n$ covariance matrix of link travel times. Since the link travel times are assumed to be independent, Σ is a diagonal matrix here with the element $\Sigma_{a,a} = \sigma_a^2$. We denote by Λ a $I \times I$ diagonal matrix with $\Lambda_{i,i} = \sum_b \delta_{i,b}\sigma_b^2$. In fact, we have the representation $\Lambda = \text{diag}(\Delta\Sigma \cdot \mathbf{1})$, where $\mathbf{1}$ is an n-by-1 vector with 1 as its element, and $\text{diag}(\cdot)$ denotes the transformation of a vector to a diagonal matrix. In this representation, the operator $\cdot$ emphasizes that the multiplication is taken between a matrix and a vector. If we let $\tilde{\Delta} = \Lambda^{-1}\Delta$, i.e., $\tilde{\Delta}$ being the incidence matrix Δ scaled by $(\sum_b \delta_{i,b}\sigma_b^2)^{-1}$ for all $\delta_{i,a}$ in the row i, Equation (5) can be written as $\tilde{\Delta}^T\Delta \cdot \mu = \tilde{\Delta}^T \cdot X$. We note that the matrix Δ is of the same rank with

the matrix $\tilde{\Delta}$. This result actually implies that a unique solution exists as long as the incidence matrix Δ is of full rank and all σ_a are known. Under this condition, Equation (5) has the solution $\mu = (\tilde{\Delta}^T\Delta)^{-1}\tilde{\Delta}^T \cdot X$, which is the weighted least squares estimation. Moreover, Equation (6) can be written as $\tilde{\Delta}^T\cdot\mathbf{1} = \Delta^T\cdot[(\Lambda^{-1}(X-\Delta\cdot\mu))\circ(\Lambda^{-1}(X-\Delta\cdot\mu))]$, where $\circ$ denotes the element-wise product.

2.1.2 Analysis of mean estimates: impact of errors in variance estimates

Equation (5) can be reduced to a series of linear equations regarding link time mean estimates, given the values of variance estimates. It can be shown that if the trip variance values are predetermined within a certain range of estimate errors, it would be computationally easy to solve for the mean link time estimates with reasonable errors. We illustrate this point below.

We let $\hat{\sigma}_b^2$ be the variance estimate used in Equation (5) and let σ_b^2 be its real value. For convenience, we assume that there is a disturbance ϵ_b in the variance estimates, i.e., $\hat{\sigma}_b^2 = \sigma_b^2 - \epsilon_b^2$ in the following analysis. A similar analysis can be applied to the case $\hat{\sigma}_b^2 = \sigma_b^2 + \epsilon_b^2$ as well as the general case $\hat{\sigma}_b^2 = (\sigma_b - \epsilon_b)^2$.

We denote by $\hat{\mu}$ the vector solution to Equation (5) with $\hat{\sigma}_b^2$. The matrix $\tilde{\Delta}$ is the same as defined before with σ_b^2, i.e., $\tilde{\Delta} = \Lambda^{-1}\Delta$ with $\Lambda_{i,i} = \sum_b \delta_{i,b}\sigma_b^2$. We also use $\|\cdot\|$ to denote the norm of matrices or vectors. Let Λ_ϵ be a $I \times I$ diagonal matrix with the i-th element in diagonal being $\frac{\sum_b \delta_{i,b}\epsilon_b^2}{\sum_b \delta_{i,b}\sigma_b^2}$, then we have the following.

Proposition 1 $\|\mu - \hat{\mu}\|$ is sufficiently small provided $\|\Lambda_\epsilon\| \ll 1$, where $\mu = (\tilde{\Delta}^T\Delta)^{-1}\tilde{\Delta}^T \cdot X$.

Proof. We have $(\sum_b \delta_{i,b}\hat{\sigma}_b^2)^{-1} = \left(\sum_b \delta_{i,b}\sigma_b^2\left(1 - \frac{\sum_b \delta_{i,b}\epsilon_b^2}{\sum_b \delta_{i,b}\sigma_b^2}\right)\right)^{-1} = (\sum_b \delta_{i,b}\sigma_b^2)^{-1}\left(1 + \frac{\sum_b \delta_{i,b}\epsilon_b^2}{\sum_b \delta_{i,b}\sigma_b^2} + ...\right)$, provided that $\frac{\sum_b \delta_{i,b}\epsilon_b^2}{\sum_b \delta_{i,b}\sigma_b^2} \ll 1$ and the higher order terms are omitted. Then the Left Hand Side (LHS) and the Right Hand Side (RHS) of Equation (5) for all links become

$$
\begin{aligned}
\text{LHS} &= \sum_i \frac{\delta_{i,a}(\sum_b \delta_{i,b}\hat{\mu}_b)}{\sum_b \delta_{i,b}\sigma_b^2} + \sum_i \frac{\delta_{i,a}(\sum_b \delta_{i,b}\hat{\mu}_b)(\sum_b \delta_{i,b}\epsilon_b^2)}{(\sum_b \delta_{i,b}\sigma_b^2)^2} + ... \ , \\
&= \tilde{\Delta}^T\Delta \cdot \hat{\mu} + (\Lambda_\epsilon\tilde{\Delta})^T\Delta \cdot \hat{\mu} + ... \ ; \qquad\qquad (7) \\
\text{RHS} &= \sum_i \frac{\delta_{i,a}x_i}{\sum_b \delta_{i,b}\sigma_b^2} + \sum_i \frac{\delta_{i,a}x_i(\sum_b \delta_{i,b}\epsilon_b^2)}{(\sum_b \delta_{i,b}\sigma_b^2)^2} + ... \ , \\
&= \tilde{\Delta}^T \cdot X + (\Lambda_\epsilon\tilde{\Delta})^T \cdot X + ... \ , \qquad\qquad (8)
\end{aligned}
$$

where the omitted terms are of a higher order of $\frac{\sum_b \delta_{i,b}\epsilon_b^2}{\sum_b \delta_{i,b}\sigma_b^2}$. Note that the second lines in Equations (7) and (8) are understood as the matrix representation for all links. Then we

have the following by omitting all higher order terms:

$$\left[\tilde{\Delta} + \Lambda_\epsilon\tilde{\Delta}\right]^T \Delta \cdot \hat{\mu} = \left[\tilde{\Delta} + \Lambda_\epsilon\tilde{\Delta}\right]^T \cdot X, \tag{9}$$

and, assuming all inverse of matrices can be performed properly [1], we have

$$\begin{aligned}\hat{\mu} &= \left(\left[\tilde{\Delta} + \Lambda_\epsilon\tilde{\Delta}\right]^T \Delta\right)^{-1}\left[\tilde{\Delta} + \Lambda_\epsilon\tilde{\Delta}\right]^T \cdot X, \\ &= (\tilde{\Delta}^T\Delta)^{-1}\tilde{\Delta}^T \cdot X - (\tilde{\Delta}^T\Delta)^{-1}(\Lambda_\epsilon\tilde{\Delta})^T\Delta(\tilde{\Delta}^T\Delta)^{-1}\tilde{\Delta}^T \cdot X + ...\end{aligned} \tag{10}$$

Since the norm of Λ_ϵ is far less than 1, i.e., $\|\Lambda_\epsilon\| \ll 1$, the norm of all matrices from the second term in Equation (10) is less than or of higher order of $\|\Lambda_\epsilon\|\|(\tilde{\Delta}^T\Delta)^{-1}\tilde{\Delta}^T \cdot X\| \ll \|\mu\|$, i.e., $\|\mu - \hat{\mu}\|$ being sufficiently small. $\square$

This proposition indicates a network property that the ratio of estimate errors to the mean link time estimates has the same order with the ratio of total errors to the trip variance estimates along a path. In other words, even if errors of some link variance estimates are relatively large, the accuracy of mean estimates is still ensured as long as the trip variance estimates have reasonable errors. This finding can help compute the mean estimates easily, by solving linear equations given that the predetermined link time variance values have reasonable errors.

2.2 Inference About Link Time Considering Unknown Path Trips

In this section, we address the estimation problem where paths of some trip observations are unknown. Finding the actual trajectory of a vehicle can be challenging especially in dense urban areas, since multiple paths may exist that are consistent with a vehicle trip observation. Given observations $D = \{x_i\}$ that consist of some trips with *labeled* (or known) paths and a portion of trips with *unlabeled* (or unknown) paths, we can divide all trip observations $D = \{x_i\}$ into two subsets: D^l represents those labeled trips, and D^u denotes those trips with unlabeled path information, i.e., $D = D^l \cup D^u$. In this case, we need to simultaneously infer the paths of recorded trips, with the objective of maximizing the total likelihood over all trip observations. For each unlabeled trip, the set of possible paths that it may traverse is predetermined in this study.

---

[1]We use $(A + B)^{-1} = A^{-1} - A^{-1}BA^{-1} + \dots$, provided $\|A^{-1}B\| < 1$ where A and B are matrices. Such inverse in Equation (10) is guaranteed by assumption of $\|\Lambda_\epsilon\| \ll 1$.

2.2.1 An algorithm for hard assignment of unknown path trips

A straightforward mechanism can be designed that iterates through two steps, following the initial assignment of unlabeled trips to paths: the first step solves for the Maximum Likelihood Estimation (MLE) of distribution parameters according to the proposed derivations for trips all with known paths; the second step adjusts the path assignment for each unlabeled trip, comparing its resulting likelihood among candidate paths based on the currently estimated parameters, and reassigning it onto the path with the maximum one.

This iterative procedure essentially applies the K-means clustering algorithm (Hastie et al., 2009), which is often used to identify clusters of data. Here we use such K-means algorithm to cluster unlabeled trips with the same OD pair based on their candidate paths. For each iteration, every unlabeled trip is assigned uniquely to a path, which may be considered as *hard* assignment in contrast to the model in the next section. However, there may be data points that lead to roughly similar likelihoods on different candidate paths. In that case, it is not clear that the hard assignment would be the most appropriate. Furthermore, the algorithm cannot guarantee the convergence. Therefore, we adopt a probabilistic approach next, known as *soft* assignment, for the unknown path trips.

2.2.2 Gaussian mixture model and EM algorithm for soft assignment of unknown path trips

We adopt a probabilistic point of view for the assignment of unlabeled trips. Instead of mapping it to a unique path, we consider each unlabeled trip to be on different possible paths with probabilities. The observed travel times of unlabeled trips with the same OD pair are thought of as a sample drawing from a multimodal distribution, where each of modality represents the random travel time on a possible path. Such view may also be known as the *soft* assignment (Bishop, 2006). Under the assumption of independent and normally distributed link times, it is natural to formulate the multimodal distribution as the classical Gaussian mixture model (Bishop, 2006; Bickel and Doksum, 2000).

The Gaussian mixture model is a parametric probability density function represented as a weighted sum of Gaussian component densities. The mixture weight for each component is usually called the mixing coefficient. In our context, the probability density for those unlabeled trip travel times with a distinct OD pair, i.e., with a distinct set of candidate paths, is a Gaussian mixture model, where every mixture component corresponds to a candidate path that has normally distributed travel times.

Our objective is to maximize the likelihood function based on the sample of observed trip

travel times as below: [2]

$$\max_{\mu,\sigma,\pi} \mathcal{LL}(\mu, \sigma, \pi \mid D) = \sum_{i \in D^l} \log \left(\mathcal{N}(x_i \mid \sum_a \delta_{i,a}\mu_a, \sum_a \delta_{i,a}\sigma_a^2) \right)$$

$$+ \sum_{i \in D^u} \log \left(\sum_{k \in K_i} \pi_k \mathcal{N}(x_i \mid \sum_a \delta_{k,a}\mu_a, \sum_a \delta_{k,a}\sigma_a^2) \right). \quad (11)$$

subject to

$$\sum_{k \in K_i} \pi_k = 1, \text{for path set } K_i \text{ with a distinct OD pair, } i \in D^u, \quad (12)$$

$$0 \le \pi_k \le 1. \quad (13)$$

where K_i denotes the set of possible paths that trip i may traverse and is predetermined in this study; μ_a and σ_a denote the estimated mean travel time and the standard deviation on link a; π_k is the mixing coefficient and the sum of all π_k for the corresponding path set is equal to 1; and $\delta_{k,a}$ by abuse of notation is equal to 1 if the path k of trip i contains the link a for $k \in K_i$, otherwise 0. Also note that $K_i = K_j$ if unlabeled trips i and j have the same OD pair.

The objective (11) leads to the following equations, by setting the partial derivative to zero for a specific link a with respect to its parameters μ_a and σ_a, respectively,

$$0 = \sum_{i \in D^l} \frac{\delta_{i,a}(x_i - \sum_b \delta_{i,b}\mu_b)}{\sum_b \delta_{i,b}\sigma_b^2} + \sum_{i \in D^u} \sum_{k \in K_i} \frac{\delta_{k,a}\gamma_k(x_i)(x_i - \sum_b \delta_{k,b}\mu_b)}{\sum_b \delta_{k,b}\sigma_b^2}, \text{ for link } a, \quad (14)$$

$$0 = \sum_{i \in D^l} \left(\frac{\delta_{i,a}}{\sum_b \delta_{i,b}\sigma_b^2} - \frac{\delta_{i,a}}{(\sum_b \delta_{i,b}\sigma_b^2)^2}(x_i - \sum_b \delta_{i,b}\mu_b)^2 \right)$$

$$+ \sum_{i \in D^u} \sum_{k \in K_i} \frac{\delta_{k,a}\gamma_k(x_i)\left[2(\sum_b \delta_{k,b}\sigma_b^2) - (x_i - \sum_b \delta_{k,b}\mu_b)^2\right]}{(\sum_b \delta_{k,b}\sigma_b^2)^2}, \text{ for link } a, \quad (15)$$

where $\gamma_k(x_i)$ represents the probability that the component (or candidate path) k takes for explaining the trip observation i:

$$\gamma_k(x_i) = \frac{\pi_k \mathcal{N}(x_i \mid \sum_a \delta_{k,a}\mu_a, \sum_a \delta_{k,a}\sigma_a^2)}{\sum_{j \in K_i} \pi_j \mathcal{N}(x_i \mid \sum_a \delta_{j,a}\mu_a, \sum_a \delta_{j,a}\sigma_a^2)}, \quad (16)$$

where $\mathcal{N}(x_i \mid \sum_a \delta_{k,a}\mu_a, \sum_a \delta_{k,a}\sigma_a^2)$ is used by abuse of notation to represent the prob-

---

[2]Note that the second term of Equation (11) essentially classifies the unknown-path trips with distinct OD pairs, and each class corresponds to a Gaussian mixture model with associated mixing coefficients to be determined. We ignore the summation over OD pairs here for the convenience of notations.

ability density function of Gaussian distribution at x_i with parameters mean $\sum_a \delta_{k,a}\mu_a$ and variance $\sum_a \delta_{k,a}\sigma_a^2$. Equation (16) provides another perspective on mixing coefficients π_k and $\gamma_k(\cdot)$. We can think of π_k as the prior probability of taking the path k for trips between a OD pair and $\gamma(\cdot)$ as the posterior probability after observing a particular trip time.

We also maximize the objective (11) with respect to the mixing coefficients π_k, taking into account the constraint (12) that requires the mixing coefficients summing up to one for unknown-path trips with a distinct OD pair. By incorporating Lagrange multipliers, we can solve for π_k by setting its partial derivative equal to zero:

$$\pi_k = \frac{\sum_{i \in D_{rs}^u} \gamma_k(x_i)}{\mid D_{rs}^u \mid}. \tag{17}$$

where D_{rs}^u denotes the set of unknown-path trips with a distinct OD pair rs.

We apply the expectation-maximization (EM) algorithm to solve for the parameter estimates, which leads to a MLE of the model if it exists. The algorithm iterates between performing an Expectation (E) step that creates a function for the expectation with respect to the latent variables (trip paths in our context) of the log-likelihood evaluated using current estimates, and a Maximization (M) step that updates the parameter estimates by maximizing the expected log-likelihood from the E-step. The detailed discussion on Gaussian mixture model and EM algorithm can be found in Dempster et al. (1977), McLachlan and Krishnan (2007), and Bickel and Doksum (2000). The EM algorithm is applied here as:

Step 1: Initialize μ_a, σ_a for all links, and mixing coefficients π_k for all mixture models (each model corresponds to unknown path trips with the same OD pair), and evaluate the initial value of the total log likelihood.

Step 2 (E-step): Evaluate the probabilities $\gamma_k(x_i)$ using the current parameter values based on Equation (16).

Step 3 (M-step): Re-estimate the parameters μ_a and σ_a sequentially using the current probabilities $\gamma_k(x_i)$: First keep current σ_a fixed, and update μ_a based on Equation (14), then update σ_a based on Equation (15).

Also update π_k accordingly: for those trips with the same OD pair, the mixing coefficients π_k are updated based on Equation (17).

Step 4: Evaluate the log likelihood as Equation (11), and check for the convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied,

return to Step 2.

It is noted that updating σ_a values in Step 3 may be challenging due to the complicated nonlinear Equation (15) in terms of travel time variances. Referring to Proposition 1 of mean estimates, for simplicity we may compute the mean estimates by following the proposed EM algorithm given the constant variances, and then update and maximize the total likelihood function to solve for variance estimates.

The proposed EM algorithm can guarantee the improvement of total log likelihood and lead to the local convergence by referring to the classic proof on the convergence of EM algorithm (Wu, 1983; Bishop, 2006).

2.2.3   Properties of the mean estimates

One question is whether Equation (14) has unique solution with known $\gamma_k(x_i)$ and σ_a in each iteration. For simplicity, we answer by only considering the case that σ_a is identical for all links. Let $\sum_b \delta_{k,b} = N_k$, $k \in K_i, i \in D^u$, denoting the number of links on the possible path k in the set K_i of unlabeled trip i. Also let $\sum_b \delta_{i,b} = N_i$, $i \in D^l$. Then Equation (14) turns to

$$
\sum_{i \in D^l} \frac{\delta_{i,a}(\sum_b \delta_{i,b}\mu_b)}{N_i} + \sum_{i \in D^u} \sum_{k \in K_i} \frac{\delta_{k,a}\gamma_k(x_i)(\sum_b \delta_{k,b}\mu_b)}{N_k}
$$
$$
= \sum_{i \in D^l} \frac{\delta_{i,a}x_i}{N_i} + \sum_{i \in D^u} \sum_{k \in K_i} \frac{\delta_{k,a}\gamma_k(x_i)x_i}{N_k}, \text{ for any link } a. \tag{18}
$$

To explain Equation (18) in the matrix form, we define an *augmented* incidence matrix Δ^* as the combination of all labeled and unlabeled trips. The observation index in Δ^* is arranged beginning with those in D^l followed by those in D^u, so that a labeled trip $i \in D^l$ corresponds to a unique row of $\delta_{i,a}$ in Δ^*, while an unlabeled trip $i \in D^u$ corresponds to multiple rows of $\delta_{k,a}$, $k \in K_i$, in Δ^* (the number of corresponding rows is the cardinality of K_i). The augmented incidence matrix Δ^* differs from the original incidence matrix for including all the possible paths for each trip in D^u. Let Δ^{**} denote a matrix after $\delta_{i,a}$ in Δ^* is scaled by $\frac{1}{N_i}$, for $i \in D^l$ and $\delta_{k,a}$ is scaled by $\frac{\gamma_k(x_i)}{N_k}$ for $k \in K_i$ and $i \in D^u$. These two matrices are illustrated as the follows.

$$\Delta^* = \begin{array}{c} \\ 1 \\ \vdots \\ k \\ k_1^1 \\ \vdots \\ k_q^j \end{array}\begin{array}{c} a_1 \qquad \cdots \qquad a_n \\ \left(\begin{array}{ccc} \delta_{1,a_1} & \cdots & \delta_{1,a_n} \\ \vdots & \ddots & \vdots \\ \delta_{k,a_1} & \cdots & \delta_{k,a_n} \\ \delta_{k_1^1,a_1} & \cdots & \delta_{k_1^1,a_n} \\ \vdots & \ddots & \vdots \\ \delta_{k_q^j,a_1} & \cdots & \delta_{k_q^j,a_n} \end{array}\right) \end{array}, \tag{19}$$

$$\Delta^{**} = \begin{array}{c} 1 \\ \vdots \\ k \\ k_1^1 \\ \vdots \\ k_q^j \end{array}\left(\begin{array}{ccc} \frac{\delta_{1,a_1}}{N_1} & \cdots & \frac{\delta_{1,a_n}}{N_1} \\ \vdots & \ddots & \vdots \\ \frac{\delta_{k,a_1}}{N_k} & \cdots & \frac{\delta_{k,a_n}}{N_k} \\ \frac{\gamma_{k_1^1}(x_{k^1})\delta_{k_1^1,a_1}}{N_{k_1^1}} & \cdots & \frac{\gamma_{k_1^1}(x_{k^1})\delta_{k_1^1,a_n}}{N_{k_1^1}} \\ \vdots & \ddots & \vdots \\ \frac{\gamma_{k_q^j}(x_{k^j})\delta_{k_q^j,a_1}}{N_{k_q^j}} & \cdots & \frac{\gamma_{k_q^j}(x_{k^j})\delta_{k_q^j,a_n}}{N_{k_q^j}} \end{array}\right), \tag{20}$$

where $1, ..., k \in D^l$ denotes the row index for each labeled trip with known path, and $k_j^i \in K_{k^i}$ denotes the row index for unlabeled trip $k^i \in D^u$ with trip time x_{k^i}. In the presentation (20) of matrix Δ^{**}, for example, $k_1^1, ..., k_m^1$ indicate that there are m possible paths for trip time x_{k^1}. In general, $k_1^j, ..., k_q^j$ indicate that there are q possible paths for the trip time x_{k^j}. Note that matrix Δ^* is of the same rank with matrix Δ^{**}.

Equation (18) is therefore rewritten as

$$(\Delta^{**})^T \Delta^* \cdot \mu = (\Delta^{**})^T \cdot X. \tag{21}$$

where X by abuse of notation denotes the column vector of trip times with proper arrangement and augmentation, i.e., x_{k^i} has $|K_{k^i}|$ duplications in X. Equation (21) and prior analysis imply the following proposition.

Proposition 2 There is a unique solution to Equation (18) provided that the augmented incidence matrix Δ^* is of full rank, and $\gamma_k(x_i)$ and σ_a are known.

2.3 Confidence Interval for Mean Link Travel Time

In practice, it is also important to obtain the confidence intervals for mean link travel times. The corresponding estimation can be approximated, for example, by the profile likelihood method, as briefly described below.

Let $\mu = (\mu_1, \ldots, \mu_n)$ denote the parameters of interest (mean link time in our context) and ϕ a vector of other parameters (i.e., nuisance parameters). Suppose we want to estimate the confidence interval for μ_1. We let $\mu_{-1} = (\mu_2, \ldots, \mu_n)$ and express the log-likelihood function as $\mathcal{LL}(\mu_1, \mu_{-1}, \phi)$. Then we may express the log-likelihood ratio statistic for parameter μ_1, denoted by $r(\mu_1)$, in terms of the profile likelihood function as

$$r(\mu_1) = 2 \left\{ \max_{\mu,\phi} \mathcal{LL}(\mu, \phi) - \max_{\mu_{-1},\phi} \mathcal{LL}(\mu_1, \mu_{-1}, \phi) \right\}. \tag{22}$$

It can be shown that $r(\mu_1)$ is asymptotically distributed as χ_1^2 (chi-square distribution with one degree of freedom) when the sample size goes to infinity (Bickel and Doksum, 2000). Therefore, the 95% confidence interval for μ_1 can be approximated as

$$\left\{ \mu_1 : r(\mu_1) \leq \chi_1^2(0.95) \right\}. \tag{23}$$

The profile likelihood method may be computationally expensive for large-scale networks. An alternative approach to estimating confidence interval is through the observed Fisher information matrix (Bickel and Doksum, 2000). In the proposed framework, the observed information matrix can be computed on the last iteration of the EM procedure. We do not present details here, but interested readers can also refer to Louis (1982).

3 METHOD II: TRIP SPLITTING APPROXIMATION

In our earlier models, Gaussian distribution of link travel times gives rise to a trip time that follows a closed form distribution, which makes modeling technically tractable. However, the link time may follow other probability distributions than the Gaussian such as the lognormal, or a mixed distribution due to the recurrent traffic congestion, in which case, no closed form distribution for trip time is available. We propose to split the trip time among traversed links. Different approaches to splitting trip travel time would lead to different estimates. In this section, we propose a statistical method of trip splitting approximation and examine its basic properties. We assume that the observations are sufficient such that the incidence matrix is of full rank, as in the Method I. We mainly focus on the case that the path of each trip observation is known and travel time on each link follows a certain

general distribution.

3.1 General Approach

We denote by D_p the set of trips traveling along path p, and the set P comprising of all paths of trips. In other words, the trips are grouped according to their paths. Let incidence indicator $\delta_{i,a}$ denote if trip i traverses link a. Trip time x_i, $i \in D_p$, actually comprises of unobserved $x_{i,a}$ on link a along path p, and hence $x_i = \sum_a \delta_{i,a} x_{i,a}$. We use $\xi_{i,a}$ to denote the corresponding random variable of travel time on link a for observed trip i, whose realized value being $x_{i,a}$, and denote by $f_a(\ \cdot\ ;\ \Theta_a)$ its probability density function with the parameter vector Θ_a. Also assume that $\xi_{i,a}$ are independent for trip i. Since the link travel time is unobserved, we have to maximize the following conditional expected log-likelihood function with respect to all parameters:

$$\mathcal{LL}(\Theta \mid D) = \sum_{p \in P} \sum_{i \in D_p} \mathbb{E} \left\{ \sum_a \delta_{i,a} \log f_a(\xi_{i,a};\ \Theta_a) \middle| \sum_a \delta_{i,a} \xi_{i,a} = x_i \right\}, \tag{24}$$

where Θ denotes the vector of all parameters in the above function. If we denote by $(x_{i,a})$ the row vector $(x_{i,a_1}, x_{i,a_2}, ..., x_{i,a_n})$ where n is number of links, and denote by $f(\cdot \mid \sum_a \delta_{i,a} \xi_{i,a} = x_i;\ \Theta) = \prod_a f_a(\cdot \mid \sum_a \delta_{i,a} \xi_{i,a} = x_i;\ \Theta_a)$ the conditional probability density function of $(\xi_{i,a})$ given the trip observation x_i, then we have

$$\mathcal{LL}(\Theta \mid D) = \sum_{p \in P} \sum_{i \in D_p} \int_{\mathbb{R}^n} f((x_{i,a}) \mid \sum_a \delta_{i,a} \xi_{i,a} = x_i;\ \Theta) \sum_a \delta_{i,a} \log f_a(x_{i,a};\ \Theta_a)\, d(x_{i,a}). \tag{25}$$

If $\delta_{i,a} = 0$ for some a in the integral in Equation (25), the corresponding $x_{i,a}$ will be automatically integrated out. Then the log-likelihood Equation (25) should be maximized according to the following

$$\mathcal{LL}(\Theta \mid D) = \sum_{p \in P} \sum_{i \in D_p} \int_{\sum_a \delta_{i,a} x_{i,a} = x_i} \frac{\prod_a f_a(x_{i,a};\ \Theta_a)}{\mathbb{P}(\sum_a \delta_{i,a} \xi_{i,a} = x_i;\ \Theta)} \sum_a \delta_{i,a} \log f_a(x_{i,a};\ \Theta_a)\, d(x_{i,a}), \tag{26}$$

where $\mathbb{P}(\sum_a \delta_{i,a} \xi_{i,a} = x_i;\ \Theta) = \int_{\sum_a \delta_{i,a} x_{i,a} = x_i} \prod_a f_a(x_{i,a};\ \Theta_a)\, d(x_{i,a})$.

The difficulty of maximizing Equation (26) is the evaluation of the multi-dimensional integral. It may be possible to employ Monte Carlo techniques to evaluate the integral, especially when the probability density enjoys some special structures. However, it generally involves expensive computation even for a small-size network, therefore it is difficult to implement in practice.

One practical approach is to approximate the conditional probability density in Equation (25) by directly splitting path travel time onto links. We assume that trips on the same path under similar traffic conditions have more or less a fixed fraction of the trip time for the same link. Let $w_{p,a}$ denote the proportion of travel time on link a among the total travel time on path p, and w be the vector of $w_{p,a}$. If the variation of the proportion $w_{p,a}$ is relatively small, the conditional probability density may be approximated by $\prod_a \delta(x_{i,a} - w_{p,a}x_i)$ where $\delta(\cdot)$ denotes the Dirac delta function (Gelfand and Shilov, 1964). This Dirac delta function notation should not be confused with the incidence notation $\delta_{i,a}$. Then the problem of maximizing Equation (25) is approximated as [3]

$$\max_{\Theta,w} \mathcal{LL}(\Theta, w \mid D) = \sum_{p \in P} \sum_{i \in D_p} \sum_a \delta_{i,a} \log f_a(w_{p,a}x_i;\ \Theta_a), \tag{27}$$

subject to

$$\sum_a w_{p,a}\delta_{p,a} = 1, \text{ for any trip path } p, \tag{28}$$

$$0 \le w_{p,a} \le 1. \tag{29}$$

In Equation (28), $\delta_{p,a}$ is used for the convenience of notations, denoting if a trip along path p traverses link a. Note that we use $\delta_{p,a}$ instead of previous notation $\delta_{i,a}$ in order to go with the notation $w_{p,a}$, and we also enforce $w_{p,a} = 0$ if $\delta_{p,a} = 0$. Similar to the EM algorithm, an iterative approach to obtain the parameters Θ can be performed by repeating the following steps until convergence. Specifically, at k-th iteration, we have

Step 1 Estimate $w_{p,a}$ by using the estimates of Θ from Step 2 in the $(k-1)$-th iteration;

Step 2 Estimate Θ by using $w_{p,a}$ obtained from Step 1.

We next consider a case with known paths of trips in which link times follow Gaussian distributions. This special case allows a comparison with Method I proposed earlier. Then we consider the case of link times following log-normal distributions.

3.2   Case of Gaussian Distribution

We assume that all link travel time variables $\xi_{i,a}$ are independent and follow Gaussian distributions, i.e., $\mathcal{N}(\cdot \mid \mu_a, \sigma_a^2)$. The objective (27) with constraints (28)–(29) leads to the following equations, by setting the partial derivative to zero for a specific link a with

---

[3]This derivation applies the property of Dirac delta function: $\int \delta(x_{i,a} - w_{p,a}x_i)g(x_{i,a})\,dx_{i,a} = g(w_{p,a}x_i)$ for any function $g(\cdot)$.

respective to its parameters μ_a and σ_a respectively,

$$\mu_a = \frac{\sum_{p\in P}\sum_{i\in D_p}\delta_{i,a}w_{p,a}x_i}{\sum_{p\in P}\sum_{i\in D_p}\delta_{i,a}}, \tag{30}$$

$$\sigma_a^2 = \frac{\sum_{p\in P}\sum_{i\in D_p}\delta_{i,a}(w_{p,a}x_i - \mu_a)^2}{\sum_{p\in P}\sum_{i\in D_p}\delta_{i,a}}, \tag{31}$$

where $\sum_{p\in P}\sum_{i\in D_p}\delta_{i,a} \neq 0$. Obviously, $\sum_{p\in P}\sum_{i\in D_p}\delta_{i,a}$ is the number of trip observations traversing link a. We maximize the objective (27) with respect to the ratio $w_{p,a}$ by considering Lagrange multipliers:

$$\sum_{p\in P}\sum_{i\in D_p}\sum_{a}\delta_{i,a}\cdot\log\left(\mathcal{N}(w_{p,a}x_i \mid \mu_a, \sigma_a{}^2)\right) + \sum_{p\in P}\lambda_p(\sum_a w_{p,a}\delta_{p,a} - 1). \tag{32}$$

Taking the partial derivative with respective to $w_{p,a}$ and solving for λ_p and $w_{p,a}$, we obtain the following equations

$$\lambda_p = \frac{\sum_{i\in D_p}x_i(x_i - \sum_a \delta_{p,a}\mu_a)}{\sum_a \delta_{p,a}\sigma_a^2}, \text{ for any trip path } p \text{ and } \sum_a \delta_{p,a}\sigma_a^2 \neq 0, \tag{33}$$

and,

$$w_{p,a} = \mu_a \cdot \frac{\sum_{i\in D_p}x_i}{\sum_{i\in D_p}x_i^2} + \lambda_p \cdot \frac{\sigma_a^2}{\sum_{i\in D_p}x_i^2}, \text{ for any trip path } p, \text{ link } a \text{ and } \delta_{p,a} \neq 0. \tag{34}$$

There is a statistical interpretation for the Lagrange multiplier. Consider a simple case where all trips have the same path p. In this case, $\sum_{i\in D_p}\delta_{i,a}$ is the same for any link a along this fixed path, and is denoted by N_p. We also denote the sum of link travel time variance σ_a^2 along the path by σ_p^2. Then plugging Equation (30) into Equation (33), we have

$$\lambda_p = \frac{\sum_{i\in D_p}x_i^2 - \frac{(\sum_{i\in D_p}x_i)^2}{N_p}}{\sigma_p^2} = \frac{s_p^2}{\sigma_p^2}(N_p - 1), \tag{35}$$

where s_p^2 denotes the sample variance of trip travel time along path p. When a large number of trips are observed, $s_p^2 \approx \sigma_p^2$, which gives rise to $\lambda_p \approx N_p - 1$.

We also note that Equation (34) cannot guarantee positive $w_{p,a}$ in some extreme cases. If this situation happens, we may either directly solve the constrained optimization (27)–(29) with fixed μ_a, σ_a^2, or re-initialize $w_{p,a}$ and then perform the iterative algorithm.

To summarize, we apply the iterative algorithm here as:

Step 1: Initialize μ_a, σ_a for all links.

Step 2: Evaluate the Lagrange multipliers λ_p using the current parameter values based on Equation (33), and update the splitting ratios $w_{p,a}$ accordingly based on Equation (34).

Step 3: Re-estimate the parameters μ_a and σ_a using the current splitting ratios $w_{p,a}$ based on Equations (30) and (31).

Step 4: Evaluate the total log likelihood as Equation (27), and check for the convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied, return to Step 1; otherwise, terminate.

3.3 Case of Log-Normal Distribution

We consider another case with known paths of trips, in which link travel time variables $\xi_{i,a}$ are independent and follow log-normal distributions for any link a of trip i. We have the probability density function $f_a(x_{i,a}; \Theta) = \frac{1}{x_{i,a}} \mathcal{N}(\log x_{i,a} \mid \mu_a, \sigma_a^2)$. The objective (27) in this case becomes

$$\mathcal{LL}(\mu, \sigma, w \mid D) = \sum_{p \in P} \sum_{i \in D_p} \sum_a \delta_{i,a} \cdot \log \left(\frac{1}{w_{p,a} x_i} \mathcal{N}(\log(w_{p,a} x_i) \mid \mu_a, \sigma_a^2) \right). \qquad (36)$$

Similarly, constraints (28)–(29) still hold.

For the estimates of parameters in Step 2 of the proposed iterative approach in Section 3.1, we have the following equations by fixing $w_{p,a}$:

$$\mu_a = \frac{\sum_{p \in P} \sum_{i \in D_p} \delta_{i,a} \log(w_{p,a} x_i)}{\sum_{p \in P} \sum_{i \in D_p} \delta_{i,a}}, \qquad (37)$$

$$\sigma_a^2 = \frac{\sum_{p \in P} \sum_{i \in D_p} \delta_{i,a} (\log(w_{p,a} x_i) - \mu_a)^2}{\sum_{p \in P} \sum_{i \in D_p} \delta_{i,a}}. \qquad (38)$$

For the estimates of $w_{p,a}$ in Step 1 in Section 3.1, there is no closed-form expression. Therefore, we can estimate $w_{p,a}$ through the nonlinear optimization (36) with constraints (28)–(29) by fixing μ_a and σ_a at each iteration.

3.4 Case with Unknown Path Trips

We briefly discuss the case that the paths of some trip observations are unknown. Similar to the previous solution framework, we can apply the EM steps at iterations to infer the

unknown paths as well as travel time estimates. If link travel time follows a Gaussian distribution, the derivation in both the E-step and the M-step may lead to a closed form. While for other general distributions such as log-normal distribution, it may not be possible to obtain closed form solutions in some steps. Taking log-normal distributed link travel time as an example, the splitting ratios $w_{p,a}$ may be derived after applying some approximations to the sum of log-normal random variables. One can also apply some optimization methods to solve for the parameters in M-step, though the computational cost would become prohibitive. In general, the EM algorithm is desirable as long as either the E-step or M-step can be solved easily. However, we leave the detailed discussions for the future work.

4 EXPERIMENTAL RESULTS

We numerically test the proposed models and procedures on networks with results being organized in three subsections. First we test the EM algorithm of Method I to find the individual algorithm efficiency, followed by testing the trip splitting approximation method for the log-normal distributed link travel times. We also compare the estimates from using both Method I and Method II with link times following Gaussian distributions. Two networks are used for the tests: a simple network with nine directional links as well as the Sioux Falls network. Each will be explained in more detail in the following sections.

All the numerical tests in this section are conducted on a Windows 7×64 Workstation with two 2.70 GHz CPUs and 4 GB RAM. We code the algorithms in MATLAB, and the convergence criterion is set that the gap of objective value of total likelihood from two consecutive iterations is no larger than 1e-4.

4.1 Test EM Algorithm for the Case with Unknown Path Trips

A simple network with nine directional links

Figure 1 shows a simple acyclic network consisting of nine directional links. All link travel times are independent and normally distributed. Trips are generated/observed to guarantee that the rank of the link-path incidence matrix has a rank equal to the number of links to estimate (i.e., a full rank system). We are aware of many ways to generate a full rank system. Here we choose a direct approach by adding single link observations. In addition, trips between two OD pairs with unknown paths are also generated: from A to F and from C to D, respectively, which means actual paths traversed by those trips are kept from the observed trips and are inferred instead by the procedure proposed earlier. The path information for trips of unknown paths is kept as the ground truth for assessing

Table 1: Illustration of Test Sample

Type of Generated Trips		Number of Trips
Trips along each link		50
Trips covering multiple links		50
Trips with unlabeled paths	From A to F	200
	From C to D	50

the parameter estimates.

We randomly generate link travel times with an arbitrary mean between 40 and 80 and a standard deviation between 6 and 20 for each link. Different times on the same link are experienced by trips, all following a normal distribution of the same mean and variance. OD trips are generated whose total travel time is the sum of the link times traversed. The generated link times are used as ground truth to assess the link estimates from the proposed methods. The test sample size is described in Table 1.
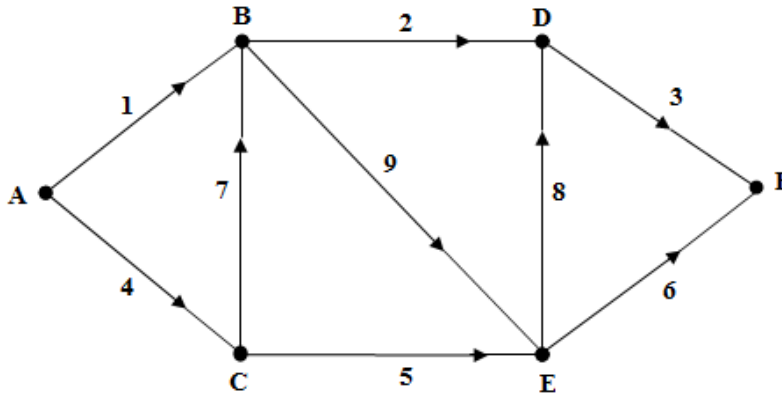

Figure 1: A simple test network.

We enumerate candidate paths of the two OD pairs whose paths are unknown. The numbers in each bracket represent the traversed links sequentially.

- From A to F: [1, 2, 3]; [4, 5, 6];

- From C to D: [5, 8]; [7, 2]; [7, 9, 8].

We then compute the estimates of link means by following the proposed EM algorithm in Method I, and solve for variance estimates by maximizing the total likelihood function.

Table 2: Estimated and Ground Truth Values of Parameters for Each Link

Link No.	Mean			Standard Deviation		
	Ground Truth	Estimate	Error	Ground Truth	Estimate	Error
1	72.6	68.4	5.78%	18.7	20.1	7.35%
2	59.6	58.1	2.39%	10.7	12.0	11.70%
3	68.9	68.8	0.09%	8.1	7.6	5.63%
4	53.6	58.4	8.84%	14.5	13.5	7.22%
5	63.0	60.6	3.80%	17.8	18.9	6.25%
6	66.7	65.9	1.19%	14.4	14.9	3.07%
7	57.0	61.2	7.50%	12.0	12.5	4.36%
8	63.3	63.7	0.58%	13.7	14.3	4.24%
9	68.9	69.9	1.50%	18.3	15.6	14.95%
MAPE	-	-	3.52%	-	-	7.20%

The variances are solved by coding the interior point method in MATLAB for the constrained nonlinear program. The total computational time to obtain estimates is about two minutes. Table 2 summarizes the resulting estimates and errors, where the Mean Absolute Percentage Error (MAPE) is recorded for each estimate. To illustrate, MAPE for the link mean estimate is calculated as

$$MAPE = \frac{1}{n} \sum_{a=1}^{n} \frac{|\mu_a - \hat{\mu}_a|}{\mu_a} \tag{39}$$

where n denotes the total number of links, $\hat{\mu}_a$ denotes the estimated mean travel time on link a, and μ_a denotes its ground truth mean value.

We also obtain the resulting mixing coefficients for each Gaussian mixture model from optimization of the total likelihood function, and their estimates from the iterative EM algorithm serve as initial guess for the nonlinear optimization. As for trips with unknown paths, the truth is that trips from A to F are equally split between the two alternative paths, and trips from C to D all traverse the path [7, 9, 8]. Table 3 shows the estimated mixing coefficients for trips on the two OD pairs, which demonstrates close proximity to the true path choices. Figure 2 shows that the proposed EM algorithm results in fast improvement of total likelihood to its convergence at iterations.

Table 3: Estimated Mixing Coefficients for Unlabeled Trips

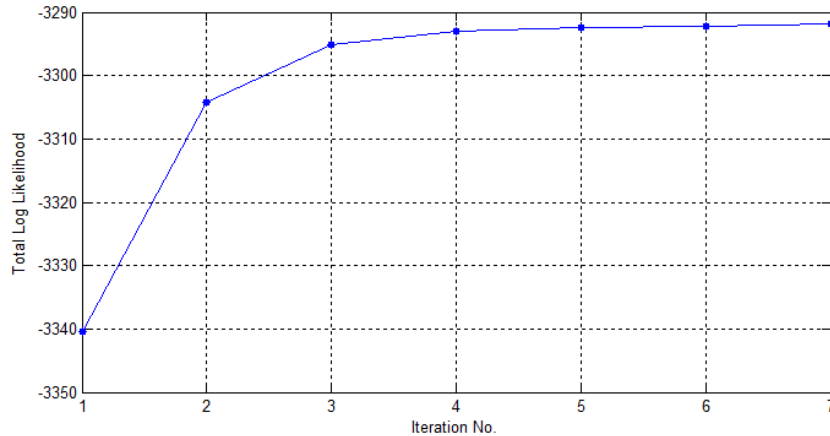| OD Pair	Candidate Paths	Mixing Coefficients	
AF	$[1, 2, 3]$	π_1^{AF}	0.5765
	$[4, 5, 6]$	π_2^{AF}	0.4235
CD	$[5, 8]$	π_1^{CD}	0
	$[7, 2]$	π_2^{CD}	0
	$[7, 9, 8]$	π_3^{CD}	1

Figure 2: The objective value of total log likelihood with iterations for EM method.

In addition, we test the effect of trip observations along individual links and the unknown-path trip observations on estimation accuracy. As shown in Table 4, the input settings are the same, except the varying number of trips along each link. We compare the resulting mean estimates using only single-link trips, trips without unlabeled ones, and all trips respectively. The results demonstrate that the proposed method for incorporating unknown-path trips can generally lead to more accurate estimates, especially when single-link trips are insufficient. In reality, it is possible to get biased estimation when using samples of single-link trips only. To illustrate this point, we modify the input setting to make the sampling of single-link observations more biased, and also double the number of trips traversing multiple links. The resulting estimates are compared in Table 4, which indicates that the proposed method can fully utilize all trip observations and make more effective improvement for the mean estimates, while the simple estimation using only single-link trips may incur larger errors.

Sioux Falls network

The Sioux Falls network, as shown in Meng and Yang (2002), consists of 76 links and

Table 4: Comparison of Mean Estimates with Basic and Modified Settings

Basic Setting				
Trips along Each Link	Percentage of Single-link Trips	MAPE for Mean Estimates		
		Estimation Using Single-link Trips	Estimation without Unlabeled Trips	Estimation Using All Trips
10	23.08%	9.11%	8.94%	6.86%
20	37.50%	8.89%	8.43%	7.18%
40	54.55%	6.64%	5.97%	5.93%
60	64.29%	4.53%	4.32%	3.44%
80	70.59%	2.43%	2.40%	2.33%
Modified Setting				
10	13.04%	16.14%	13.93%	10.14%
20	23.08%	13.69%	12.77%	8.33%
40	37.50%	13.63%	12.46%	8.19%
60	47.37%	10.44%	10.15%	7.57%
80	54.55%	6.27%	6.19%	5.06%

24 nodes. Based on the predetermined Gaussian distribution on each link, we randomly generate the sample of trip travel times. Trips with 'unknown' paths are also generated similarly as for the nine-link network earlier. The trips generated again guarantees a full rank system. The input information for the estimation analysis is summarized in Table 5.
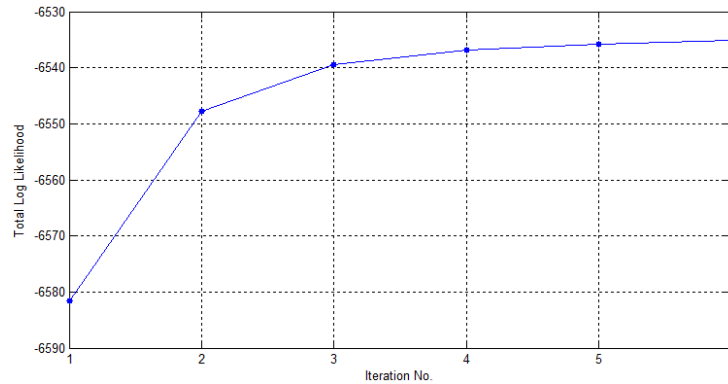

Figure 3: The objective value of total log likelihood with iterations for the EM method.

Table 6 illustrates estimation errors using the Sioux Falls network, which appear to be within range of general acceptance. Figure 3 shows that for the Sioux Falls network

Table 5: Basic Input Information to Generate Test Sample for the Case with Unknown Path Trips

Type of Generated Trips		Number of Trips
Trips along each link		10
Trips covering multiple links		550
Trips with unlabeled paths		300
Setting of Randomly Generated Parameters		Value of Bounds
Mean	Upper Bound	70
	Lower Bound	40
Standard Deviation	Upper Bound	20
	Lower Bound	6

Table 6: Estimate Errors for All Links

Estimated Parameters	MAPE
Mean	4.68%
Standard Deviation	12.16%

the proposed EM algorithm can still lead to convergence of the total likelihood within 6 iterations.

As noted earlier, the application of a nonlinear solver in MATLAB may experience computational issues due to the large number of variables. The total computational time here is nearly 20 minutes. Besides, its ability to search for good solutions appears challenged. Therefore, the design of heuristic algorithms for this particular constrained nonlinear problem is meaningful for future studies.

To further examine the computational performance of proposed EM algorithm, we test the same Sioux Falls network but with varying number of unknown-path trips as illustrated in Table 7. Figure 4 displays that the total computational time increases fast with a larger sample of unknown-path trips.

Table 7: Computational Time of the EM method for a Varying Number of Unknown-path Trips

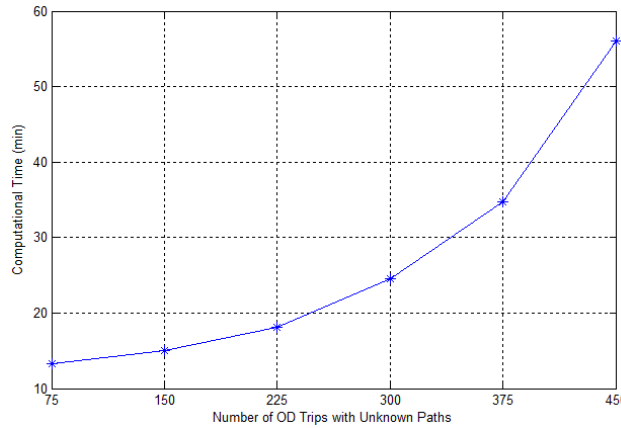| Input Setting for Unknown-path Trips		Computational Time (min)
OD Pairs / Trips Per Pair	Total Number of Trips	
5 / 15	75	13.26
5 / 30	150	15.00
5 / 45	225	18.15
5 / 60	300	24.49
5 / 75	375	34.69
5 / 90	450	56.05

Figure 4: Computational time with a varying number of unknown-path trips.

4.2 Test Trip Splitting Approximation for the Case of Log-Normal Distribution

A simple network with nine directional links

In this case, all the link travel times are log-normally distributed. We generate link times with randomly selected mean values between 40 and 80, and standard deviation values between 6 and 20, as before. The ground truth of parameters for each link are the same as listed in Table 2. Specifically, we sample both the trips along a single link and trips covering multiple links as illustrated in Table 8, where the numbers in each bracket represent the traversed links sequentially, and we randomly generate fifty trips on each path.

We solve for the parameter estimates for link travel times using trip splitting approxi-

Table 8: Illustration of Generated Trip Itineraries

Path No.	1	2	3	4	5	6
Link Sequence	[1,2,3]	[1,9,8,3]	[4,5,6]	[5,6]	[7,9,8]	[4,7,9]

Table 9: Estimate Errors for Each Link

Link No.	Estimate Errors			
	Mean	SD	Mu	Sigma
1	0.60%	13.01%	0.33%	13.19%
2	1.61%	1.94%	0.40%	0.34%
3	1.84%	8.03%	0.41%	6.03%
4	3.22%	23.68%	1.20%	25.47%
5	5.88%	25.53%	1.13%	20.31%
6	2.44%	12.07%	0.49%	9.68%
7	4.56%	2.68%	1.18%	6.79%
8	0.12%	16.67%	0.20%	16.48%
9	4.40%	24.66%	1.41%	27.25%
MAPE	2.74%	14.25%	0.75%	13.95%

mation. It takes less than one minute to obtain the estimates. The errors of resulting estimates are displayed in Table 9, where Mean and SD denote the mean and standard deviation of travel times on each link following log-normal distribution, while Mu and Sigma denote the parameters of the corresponding normal distribution.

We also find that the splitting ratios between the same pair of links for different trips are consistent. The results are illustrated in Table 10, where $w_{p,a}$ denotes the resulting proportion of travel time on link a among those trips along path p. Besides, Figure 5 shows that the trip splitting approximation results in fast improvement of total likelihood with iterations to convergence.

Table 10: Comparison of Splitting Ratios between a Same Link Pair along Various Paths

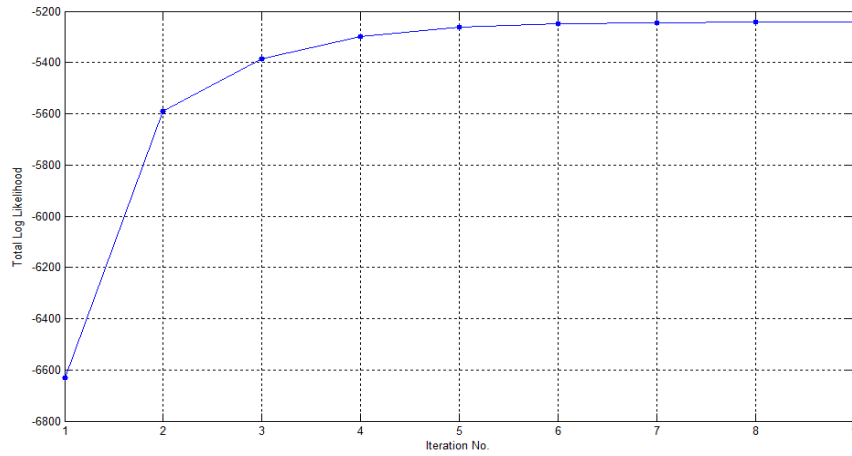| Link Pairs		Notations	Results
Links 1 and 3	Path 1	$w_{1,1}/w_{1,3}$	1.0543
	Path 2	$w_{2,1}/w_{2,3}$	1.0559
Links 5 and 6	Path 3	$w_{3,5}/w_{3,6}$	0.9018
	Path 4	$w_{4,5}/w_{4,6}$	0.9036
Links 7 and 9	Path 5	$w_{5,7}/w_{5,9}$	0.8223
	Path 6	$w_{6,7}/w_{6,9}$	0.8157
Links 8 and 9	Path 2	$w_{2,8}/w_{2,9}$	0.8724
	Path 5	$w_{5,8}/w_{5,9}$	0.8793

Figure 5: The objective value of total log likelihood with iterations for trip splitting approximation.

Then, we test the effect of standard deviations of link travel times on the estimation accuracy using trip splitting approximation. Table 11 shows that the errors of mean estimates increase with the larger standard deviations of link times, where Mean, SD and Mu are defined the same as in Table 9. It is worth noting that the application of trip splitting approximation would have an issue as the link travel times become more unstable (i.e., with particularly large standard deviation). This can be explained by its underlying assumption of relatively stable traffic conditions on the network. For the practical applications, those individual links with heavy congestion or unexpected incidents need to be identified and carefully examined, which is beyond the scope of this study.

Sioux Falls network

We additionally test log-normal distributions for links on the Sioux Falls network. The

Table 11: Comparison of Mean Estimates with Varying Standard Deviations

Interval to Generate Random Numbers for SD	MAPE for Estimates	
	Mean	Mu
[3, 6]	0.91%	0.23%
[6, 20]	2.74%	0.75%
[20, 30]	4.97%	1.40%
[30, 40]	7.25%	2.06%

Table 12: Basic Input Information to Generate Test Sample for the Case of Log-Normal Distribution

Type of Generated Trips		Number of Trips
Trips along each arc		10
Trips covering multiple links		810
Setting of Randomly Generated Parameters		Value of Bounds
Mean	Upper Bound	70
	Lower Bound	40
Standard Deviation	Upper Bound	20
	Lower Bound	6

basic input information is summarized in Table 12. Note that all trips are with labeled paths. Table 13 illustrates the estimation errors from the trip splitting approximation, where Mean, SD, Mu, and Sigma are defined the same as in Table 9. Figure 6 indicates a fast convergence of the total likelihood with iterations.

Table 13: Estimate Errors for All Links

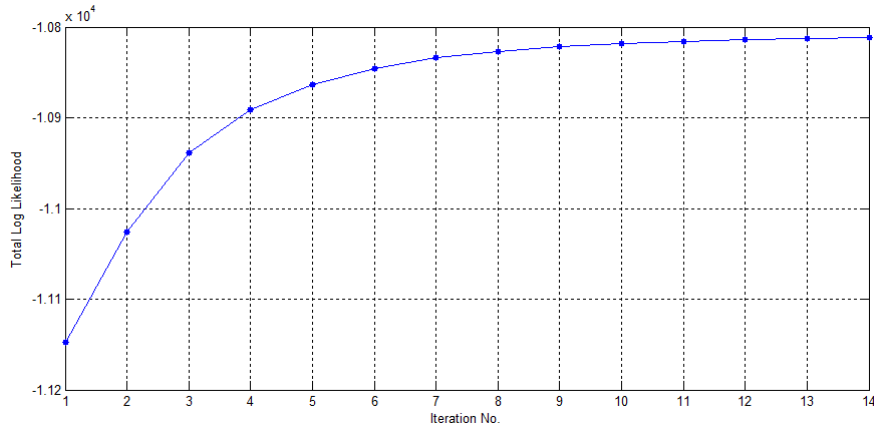| Estimated Parameters	MAPE
Mean	5.48%
SD	15.73%
Mu	1.40%
Sigma	14.98%

Figure 6: The objective value of total log likelihood with iterations for trip splitting approximation.

In the case of log-normal distribution, solving for splitting ratios in the nonlinear optimization can be computationally expensive at iterations. The total computational time is about 15 minutes in this example. A good initial point of splitting ratios is important. In practice we can always refer to the travel speeds or distances along consecutive links to provide a good starting point of splitting ratios.

4.3 Compare the Estimates Using Two Methods for the Case of Gaussian Distribution

In the case of Gaussian distributions for link travel times, we compare the estimates from using both Method I and Method II as they apply to the simple nine-link network and the Sioux Falls network respectively. The test sample size is the same as used in Section 4.2 and all trips covering multiple links are still with labeled paths. Tables 14 and 15 summarize the estimates from both methods. We also calculate the 95% confidence interval of the mean estimates for several links on the nine-link network, as illustrated in Table 16. The resulting confidence intervals for mean estimates appear very close under the two model approaches.

The trip splitting approximation of Method II generally runs very fast at iterations compared with Method I in the case of Gaussian distribution. For example, it takes only a few seconds for Sioux Falls network using Method II, as compared with nearly 3 minutes using Method I. This is likely because Method I takes a relatively long time to solve for variance estimates in the nonlinear optimization. However, the estimation accuracy of trip splitting approximation may be affected, especially that of the variance estimates.

Table 14: Comparison of Estimate Errors Using Both Methods on nine-link Network

Link No.	Estimate Errors			
	Mean		Standard Deviation	
	Method I	Method II	Method I	Method II
1	0.27%	0.68%	12.01%	14.44%
2	1.85%	1.85%	6.24%	2.21%
3	0.65%	1.50%	3.35%	1.05%
4	4.57%	4.16%	11.70%	25.66%
5	5.00%	4.93%	4.77%	20.77%
6	3.45%	3.30%	2.05%	10.60%
7	3.73%	4.22%	10.31%	7.42%
8	0.40%	1.51%	0.85%	17.12%
9	4.63%	4.09%	7.30%	28.84%
MAPE	**2.73%**	**2.92%**	**6.51%**	**14.23%**

Table 15: Comparison of Estimate Errors Using Both Methods on the Sioux Falls Network

Estimated Parameters	MAPE	
	Method I	Method II
Mean	4.98%	5.38%
Standard Deviation	9.00%	15.59%

Table 16: Illustration of 95% Confidence Interval (CI) Calculation for nine-link Network

Link No.	Method I		Method II	
	Mean Estimate	CI	Mean Estimate	CI
1	72.79	[69.20, 76.99]	73.08	[70.30, 76.49]
4	56.10	[53.38, 58.61]	55.88	[53.73, 58.12]
8	63.59	[60.75, 66.43]	63.67	[61.26, 65.91]

To summarize the trade off, the trip splitting approximation may incur larger errors with estimating the standard deviation of link travel time in trade for a much faster time compared with Method I. In terms of the mean estimates, both methods are comparably competitive.

5 DISCUSSION OF THE TWO METHODS

The two proposed methods are based on the additive property of link time distributions, i.e., whether the summation of link travel times has a closed-form distribution. Starting from the likelihood principle, both methods, whenever necessary, decompose the link travel time inference into structural steps that share the same spirit of the EM machinery. The key strategy is the introduction of the augmented data (or complete data), namely augmenting the observed data with hidden (unobserved) variables that represent the problem structure.

In Method I, the unobserved variables represent the path choices for individual travelers with unknown paths. While the proposed method involves path inference, it mainly focuses on the estimation of model parameters (so as to approximate the real values) and the stable solution, rather than the accuracy of individual path inference. The investigation of the case with all trips of known paths reveals its connection to a least squares solution. And the analysis of the property of mean estimates when trip variance estimates are within reasonable errors demonstrates the validity of our iterative calculation of mean and variance. The hard-assignment algorithm that addresses the case with some trips of unknown paths usually provides the initial solution to the soft-assignment algorithm. When dealing with the uncertainty of path choices, applying the mixture model in the soft assignment is more appropriate.

While similar methods to Method I have somewhat been studied in literature, the proposed Method II appears new. The method of splitting trip travel time is straightforward, but directly applying it cannot guarantee certain properties of results. We show that this method can be viewed from the statistical perspective, and redesign the method through maximum conditional likelihood function. The trip splitting approximation is faster in computation compared to Method I, and can be applied to various link time distributions. But it requires many parameters. Some variable selection techniques (Fan et al., 2014) may be used to overcome the proliferation of parameters. Moreover, since the E-step involves a probability inference for the augmented variables based on observed data, properly defining the augmented variables can help improve the convergence of the algorithm (Meng and Van Dyk, 1997). The proposed method of trip splitting approximation is built mainly from the statistical perspective, which can further combine the results from

conventional traffic flow theory in order to obtain more reliable estimates for practical applications. For example, one may incorporate the empirical speed-volume relationship into the iterative procedure to generate reasonable splitting ratios.

The maximum-likelihood model framework can also be extended to deal with the correlation among different links. Our modeling approaches will lay a methodological foundation that we can use to extend to a dynamic network with time-dependent link travel times (e.g., Xing et al. (2013)), to the OD flow estimation (e.g., Parry and Hazelton (2012)), to the travel time reliability (e.g., Ng et al. (2011)), or to the day-to-day dynamic travel pattern inference (e.g., Parry and Hazelton (2013)), as future study efforts. As different sources of traffic data become available (Zheng et al., 2014; Mori et al., 2014), we will find the proposed methods useful.

6 CONCLUSIONS

Link travel time estimation on a roadway network is essential for performance assessment in order to improve traffic mobility and network efficiency. It is made possible now by the widely available traffic data. This paper develops two model frameworks based on statistical inference methods for link travel time estimation using entry/exit information of trips on a network.

First, we propose a method using independent Gaussian link travel times. We particularly analyze the property of mean estimates and investigate the uniqueness of solutions. Two cases are discussed in detail: one with known paths of all trips and the other with unknown paths of some trips. We apply the Gaussian mixture model and the EM algorithm to deal with the latter. To overcome the modeling challenge that random link times do not typically add up to a trip time with close-form distribution, we develop a trip splitting approximation method assuming a relatively reliable way to partition the trip time between links. The proposed trip splitting approximation applies to the general case with arbitrary link travel time distribution, although with varying complexity in computation, making it potentially applicable to many traffic situations. The proposed methods are tested and compared numerically on two networks, a simple nine-link network and the Sioux Falls network. The experimental results indicate that both methods perform well and generate quality estimates, and that the trip splitting approximation generally runs much faster. A trade off is that the trip splitting approximation incurs relatively larger errors for standard deviation estimates than the first method.

Worthy of a special mention is that link travel time inference on a network is complicated and much still remains to be addressed. For example, can we obtain reliable link travel

time estimates if the mapping relationship between trip paths and link travel times has the rank deficiency issue? How do we improve the computational performance when solving for large number of estimates for a large-size network? Besides, further extension to the Bayesian approach with more traffic data available as prior information is also worth our examinations, and the application to various realistic networks with empirical data would be desirable in our future work.

ACKNOWLEDGEMENTS

References

Hillel Bar-Gera. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from israel. *Transportation Research Part C: Emerging Technologies*, 15(6):380–391, 2007.

Robert L. Bertini and Sutti Tantiyanugulchai. Transit buses as traffic probes: empirical evaluation using geo-location data. *Transportation Research Record: Journal of the Transportation Research Board*, 1870:35–45, 2004.

Peter J. Bickel and Kjell A. Doksum. *Mathematical statistics: basic ideas and selected topics.* Prentice Hall, 2000.

Christopher M. Bishop. *Pattern recognition and machine learning.* Springer New York, 2006.

K. S. Chan, William H. K. Lam, and Mei Lam Tam. Real-time estimation of arterial travel times with spatial travel time covariance relationships. *Transportation Research Record: Journal of the Transportation Research Board*, 2121(1):102–109, 2009.

Shyang-Lih Chang, Li-Shien Chen, Yun-Chung Chung, and Sei-Wan Chen. Automatic license plate recognition. *Intelligent Transportation Systems, IEEE Transactions on*, 5 (1):42–53, 2004.

Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from

incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

Jianqing Fan, Fang Han, and Han Liu. Challenges of big data analysis. *National Science Review*, 1:293–314, 2014.

Israel M Gelfand and GE Shilov. *Generalized functions. Vol. I: Properties and operations, Translated by Eugene Saletan.* Academic Press, New York, 1964.

Ali Haghani, Masoud Hamedi, Kaveh Farokhi Sadabadi, Stanley Young, and Philip Tarnoff. Data collection of freeway travel time ground truth with bluetooth sensors. *Transportation Research Record: Journal of the Transportation Research Board*, 2160 (1):60–68, 2010.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*, volume 2. Springer, 2009.

Bruce Hellinga, Pedram Izadpanah, Hiroyuki Takada, and Liping Fu. Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. *Transportation Research Part C: Emerging Technologies*, 16(6):768–782, 2008.

Aude Hofleitner, Ryan Herring, Pieter Abbeel, and Alexandre Bayen. Learning the dynamics of arterial traffic from probe data using a dynamic bayesian network. *IEEE Transactions on Intelligent Transportation Systems*, 13(4):1679–1693, 2012a.

Aude Hofleitner, Ryan Herring, and Alexandre Bayen. Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning. *Transportation Research Part B: Methodological*, 46(9):1097–1122, 2012b.

Timothy Hunter, Ryan Herring, Pieter Abbeel, and Alexandre Bayen. Path and travel time inference from gps probe vehicle data. *NIPS Analyzing Networks and Learning with Graphs*, 2009.

Erik Jenelius and Haris N. Koutsopoulos. Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B: Methodological*, 53:64–81, 2013.

Thomas A Louis. Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 226–233, 1982.

Xiaolei Ma, Yao-Jan Wu, Yinhai Wang, Feng Chen, and Jianfeng Liu. Mining smart card data for transit riders travel patterns. *Transportation Research Part C: Emerging Technologies*, 36:1–12, 2013.

Geoffrey McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.

Qiang Meng and Hai Yang. Benefit distribution and equity in road network design. *Transportation Research Part B: Methodological*, 36(1):19–35, 2002.

Xiao-Li Meng and David Van Dyk. The em algorithman old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59 (3):511–567, 1997.

Usue Mori, Alexander Mendiburu, Maite Álvarez, and Jose A. Lozano. A review of travel time estimation and forecasting for advanced traveller information systems. *Transportmetrica A: Transport Science*, (ahead-of-print):1–39, 2014.

ManWo Ng, W.Y. Szeto, and S Travis Waller. Distribution-free travel time reliability assessment with probability inequalities. *Transportation Research Part B: Methodological*, 45(6):852–866, 2011.

Katharina Parry and Martin L Hazelton. Estimation of origin–destination matrices from link counts and sporadic routing data. *Transportation Research Part B: Methodological*, 46(1):175–188, 2012.

Katharina Parry and Martin L Hazelton. Bayesian inference for day-to-day dynamic traffic models. *Transportation Research Part B: Methodological*, 50:104–115, 2013.

Hesham Rakha, Ihab El-Shawarby, Mazen Arafeh, and Francois Dion. Estimating path travel-time reliability. In *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, pages 236–241. IEEE, 2006.

Treerapot Siripirote, Agachai Sumalee, David P Watling, and Hu Shao. Updating of travel behavior model parameters and estimation of vehicle trip chain based on plate scanning. *Journal of Intelligent Transportation Systems*, 18:393–409, 2013.

Yilun Wang, Yu Zheng, and Yexiang Xue. Travel time estimation of a path using sparse trajectories. In *Proceeding of the 20th SIGKDD conference on Knowledge Discovery and Data Mining*, 2014.

Tao Wen, Chang Cheng Li, Chun Jiang Che, Lian De Zhong, and Xin Xin. Research on link travel time reliability model based on massive expressway toll data. *Applied Mechanics and Materials*, 505:719–726, 2014.

Bradford S. Westgate, Dawn B. Woodard, David S. Matteson, and Shane G. Henderson. Travel time estimation for ambulances using bayesian data augmentation. *The Annals of Applied Statistics*, 7(2):1139–1161, 2013.

Bradford S Westgate, Dawn B Woodard, David S Matteson, and Shane G Henderson. Large-network travel time distribution estimation for ambulances. *Submitted to European Journal of Operational Research*, 2014.

C. F. Jeff Wu. On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95–103, 1983.

Tao Xing, Xuesong Zhou, and Jeffrey Taylor. Designing heterogeneous sensor networks for estimating and predicting path travel time dynamics: An information-theoretic modeling approach. *Transportation Research Part B: Methodological*, 57:66–90, 2013.

Jean-Luc Ygnace, Chris Drane, Y. B. Yim, and Renaud De Lacvivier. Travel time estimation on the san francisco bay area network using cellular phones as probes. *California Partners for Advanced Transit and Highways (PATH) UCB-ITS-PWP-2000-18*, September 2000.

Fangfang Zheng and Henk Van Zuylen. Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, 31:145–157, 2013.

Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: concepts, methodologies, and applications. *ACM Transaction on Intelligent Systems and Technology (ACM TIST)*, 2014.

A   APPENDIX: Alternative representation of Equations (5) and (6)

We present another way to format Equations (5) and (6) by the path-link relationship. The same approach is applied to other equations in this paper. Let the path-link matrix be $\Delta_P = [\delta_{p_i,a}]_{v \times n}$, where v is the number of all paths, p_i the i-th path connecting an origin and a destination, and n the total number of links. We also denote by n_{p_i} the number of observations and by $\bar{x}_{p_i}$ the average of the travel time along path p_i. Then

Equations (5) and (6) are read as

$$0 \;=\; \sum_{p_i} \frac{\delta_{p_i,a} n_{p_i} \left(\sum_b \delta_{p_i,b} \mu_b - \bar{x}_{p_i} \right)}{\sum_b \delta_{p_i,b} \sigma_b^2}, \text{ for link } a, \tag{40}$$

$$0 \;=\; \sum_{p_i} \frac{\delta_{p_i,a} n_{p_i}}{\left(\sum_b \delta_{p_i,b} \sigma_b^2 \right)^2} \left(\sum_b \delta_{p_i,b} \sigma_b^2 - \frac{1}{n_{p_i}} \sum_{j \in p_i} \left(x_j - \sum_b \delta_{p_i,b} \mu_b \right)^2 \right), \text{ for link } a. \tag{41}$$

where $j \in p_i$ means that the observation j associates with path p_i. Let $\tilde{\Delta}_P$ be the matrix Δ_P scaled by $n_{p_i} \left(\sum_b \delta_{p_i,b} \sigma_b^2 \right)^{-1}$ for all $\delta_{p_i,a}$ in the row p_i, and let $\bar{X}$ be the vector of $\bar{x}_{p_i}$. Then Equation (40) can be also written as

$$\tilde{\Delta}_P^T \Delta_P \cdot \mu = \tilde{\Delta}_P^T \cdot \bar{X}. \tag{42}$$

Comparing with the matrix representation in Section 2.1.1, the above equation is more compact and saves memory for numerical computation. However, it seems that Equation (41) does not have a more compact representation than the one in Section 2.1.1.