View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Loughborough University Institutional Repository

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

creative commons

COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license).

Disclaimer

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

Markov Models for Bayesian Analysis about Transit Route Origin-Destination Matrices

Baibing Li
Business School
Loughborough University
Loughborough, LE11 3TU, United Kingdom

(email: b.li2@lboro.ac.uk)

Abstract

The key factor that complicates statistical inference for an origin-destination (O-D) matrix is that the problem *per se* is usually highly underspecified, with a large number of unknown entries but many fewer observations available for the estimation. In this paper, we investigate statistical inference for a transit route O-D matrix using on-off counts of passengers. A Markov chain model is incorporated to capture the relationships between the entries of the transit route matrix, and to reduce the total number of unknown parameters. A Bayesian analysis is then performed to draw inference about the unknown parameters of the Markov model. Unlike many existing methods that rely on iterative algorithms, this new approach leads to a closed-form solution and is computationally more efficient. The relationship between this method and the maximum entropy approach is also investigated.

Keywords: Bayesian analysis; Markov model; Maximum entropy method; O-D matrix; Transit route.

1. Introduction

An origin-destination (O-D) matrix provides fundamental information on flows of vehicles or people traveling from one specific geographical area to another. It plays a crucial role in traffic and transportation management.

In practice, large-scale direct sampling for statistical inference for O-D matrices is usually too expensive (Van Zuylen and Willumsen 1980; Li, 2005; Li and Cassidy, 2007). A common approach for the estimation of an O-D matrix is to calculate its entries using traffic counts obtained on pre-selected links of a transport network, without imposing any specific model on the entries. This approach is much cheaper than the method of large-scale sampling. For economic reasons, however, the number of selected links is relatively small in practice, resulting in a highly underspecified problem with a huge number of unknown parameters (entries) but many fewer observations (see e.g., Van Zuylen and Willumsen 1980). Furthermore, in some applications, even were counts to be observed on every link, the O-D matrix would still be underspecified (Hazelton, 2001; Li 2005).

In order to deal with this underspecified problem, Van Zuylen and Willumsen (1980) proposed a maximum entropy method that determines an O-D matrix such that the chosen trip matrix adds as little information as possible to the knowledge contained in the data collected on pre-selected links of a transport network.

On the other hand, many existing approaches assume that prior knowledge about an O-D matrix is more or less available (e.g., obsolete O-D matrices). An early approach, termed the balancing method, relies on a *reference* (also called *seed*) trip matrix that is composed of initial estimates obtained from a previous year. On the basis of observed totals of trips for each row and column, this O-D matrix is adjusted by multiplying each entry in a matrix row by a constant such that the row's total matches the actual count, and repeating this for each column. This iterative process continues to convergence (Lamond and Stewart, 1981). There are two

major issues associated with the balancing method. First, from a computational perspective, the balancing method is subject to the problem of non-structural zeros (Ben-Akiva et al., 1985), i.e., if an entry in a reference matrix is zero, then this entry retains a zero in every iteration. Further, when the number of zero entries becomes large, this algorithm may even fail to converge. Secondly, from a statistical point of view, this method does not incorporate an efficient way to make use of prior information because it fails to properly take into account the relative precision of prior information compared to that of current observations.

A much better approach in terms of utilizing prior information was developed by Maher (1983) and Hazelton (2001), who carried out Bayesian analysis to combine prior information with current observations on traffic flow. Recently, Li (2005) has developed a general approach to draw inference about O-D matrices; this approach can be nicely linked to many existing methods, such as the Bayesian approaches of Maher (1983) and Hazelton (2001), and the maximum entropy method of Van Zuylen and Willumsen (1980).

In this paper, we consider a particular type of problem: statistical inference for O-D matrices of transit-route ridership using on-off counts of passengers. In contrast to a general O-D matrix, a transit route O-D matrix possesses some important characteristics. So instead of estimating the entries of the O-D matrix as individual parameters, we will use these characteristics to build a parsimonious model with a much smaller number of parameters.

A parametric approach for the estimation of transit route O-D matrices was first incorporated by Li and Cassidy (2007), where all the stops of a bus route were classified into two categories, major and minor stops. Then the conditional probability that a passenger alights at a major (or minor) stop given that the passenger boarded at a major (or minor) stop was modeled and estimated using on-off counts of passengers. The entries of an O-D matrix were calculated on the basis of these conditional probabilities. This approach was shown to have many computational advantages over the balancing method (Li and Cassidy, 2007).

This paper also follows a parametric approach for the estimation of transit route O-D matrices. However, instead of classifying stops into just two categories (major and minor), this paper incorporates a Markov chain model to describe the relationships between entries of a transit route matrix. Consequently, the entries of the O-D matrix are characterized by a small number of parameters in the model, i.e., Markov transition probabilities.

To draw inference about the parameters in the Markov model, we carry out a Bayesian analysis to combine prior information with the current on-off counts of passengers. In comparison with the balancing method, this approach provides an efficient way to utilize prior information. Furthermore it leads to a closed-form solution, so its computational cost is much lower than that of existing methods. In the special case where no prior knowledge is available in practice, the prior in the Bayesian analysis can simply be chosen as non-informative so that the statistical inference is based solely on observed on-off counts.

Further, we show that when no prior is available, the developed method produces an estimate of a trip matrix that is equivalent to that obtained by the well-known maximum entropy approach. For transit route O-D estimation problems, this suggests a very close link between the maximum entropy method and the implicit assumption on the 'forgetfulness' in the Markov model, which is easily overlooked in practice when applying the maximum entropy method in the scenario of no prior information.

Finally we note that for some applications the first-order Markov process may not be able to model the reality well. Hence in this paper we also investigate Bayesian analysis for transit route O-D estimation using a higher order Markov model. With a reasonably large order m, an mth-order Markov model can provide a satisfactory approximation to the reality.

The paper is organized as follows. In Section 2 we develop a Markov model for the estimation of transit route matrices. Section 3 is devoted to Bayesian inference for the parameters in the Markov model. In Section 4 we investigate the relationship between the

proposed method and the maximum entropy approach. The obtained results are then extended to a more general situation in Section 5. In Section 6 the proposed method is illustrated using a practical example. Finally, concluding remarks are given in Section 7.

2. A Markov Model for Transit Route Matrices

Consider a (bus) route serving N stops at which transit passengers board or alight. In this section we investigate a model for statistical inference for a transit route matrix using on-off counts of passengers.

2.1. Counts of Passengers

Let y_i be the observed number of passengers who board at stop i (i=1,…,N) and let z_j be the observed number of passengers who alight at stop j (j=1,…,N), where it is assumed that no passengers board at the terminal stop, i.e. $y_N \equiv 0$, and no passengers alight at the initial stop, i.e. $z_1 \equiv 0$. Define x_{ij} to be the *unobservable* counts of passengers boarding at stop i and alighting at stop j (i, j=1,…,N). Finally, define p_{ij} to be the probabilities of passengers alighting at stop j given that they boarded at stop i (i, j=1,…,N). Figure 1 illustrates a transit route. Due to the nature of transit routes, we have

$$p_{ij} = 0 \quad \text{and} \quad x_{ij} = 0 \quad \text{for all } i \geq j .$$

(Figure 1 is about here)

Figure 1. A diagram of a bus transit route serving N stops, where x_{ij} are counts of passengers boarding at stop i and alighting at stop j, y_i and z_i are on and off counts of passengers at stop i respectively.

The counts of passengers are linked through the following equations:

$$\sum_{j=i+1}^{N} x_{ij} = y_i \qquad (i=1,\ldots,N\text{-}1), \tag{1}$$

and

$$\sum_{i=1}^{j-1} x_{ij} = z_j \qquad (j=2,\ldots,N), \tag{2}$$

as displayed in Table 1.

(Table 1 is about here)

2.2. A Markov Chain Model

Following Hazelton (2001) and Li (2005), we distinguish two different problems for transit routes: the estimation problem and the reconstruction problem. The aim for the former is to estimate unknown parameters, e.g. an alighting probability matrix $\mathbf{P} = [p_{ij}]_{N\times N}$, whereas the aim for the latter is to reconstruct actual numbers of passengers, i.e., a trip matrix $\mathbf{X} = [x_{ij}]_{N\times N}$ occurred during the observational time period. A major advantage of drawing inference about the alighting probability matrix rather than the trip matrix *per se* is that the probability parameters are more likely to remain constant across transit trips made under similar conditions (Li and Cassidy, 2007). The reconstruction of $\mathbf{X} = [x_{ij}]$ is straightforward once the alighting probability matrix **P** has been estimated. So we focus on the former problem in this paper.

In general, for the estimation problem for O-D matrices, the number of unknown quantities is much larger than the number of observations except for some special scenarios such as the estimation of intersection O-D matrices (see e.g., Li and De Moor, 1999 and 2002). When the relationships between the entries of an O-D matrix are ignored so that the entries have to be estimated individually, this is an underspecified problem. To circumvent this

difficulty, a parametric model will be developed in this paper to reduce the number of unknown parameters.

To specify an appropriate model, we first note that due to the nature of transit routes, the transition probability that a passenger will alight at stop *i* given that he/she is on board at stop $i-1$ is crucial. Consequently, rather than estimating individual entries of a transit route matrix directly, we use a first-order Markov model (termed Markov model hereafter) to characterize transition probabilities.

A Markov chain is a stochastic process where given the present state, the past and future states are independent. In a Markov model, transition probabilities play an important role. Let ξ_i be a random variable representing a passenger's state at stop *i*, where $\xi_i = 1$ if the passenger is on board at stop *i* and $\xi_i = 0$ otherwise. The Markov transition probabilities of the Markov model for a transit route are defined by:

$$\Pr\{\xi_i = k \mid \xi_{i-1} = m\} = \begin{cases} q_i & \text{if } k = 0 \text{ and } m = 1 \\ 1 - q_i & \text{if } k = m = 1 \end{cases} \qquad \text{for } i=2,\ldots,N\text{-}1, \tag{3}$$

where q_i is the probability that a passenger will alight at stop *i* given that he/she is on board at stop $i-1$. Clearly, we have $q_N \equiv 1$.

Using the Markov transition probabilities, the alighting probability matrix $\mathbf{P} = [p_{ij}]$ can be calculated as follows. First, at stop 1, we have $p_{12} = q_2$ by definition. Next, using the properties of Markov chains, we have

$$p_{13} = \Pr\{\text{alighting at stop}3 \mid \text{boarding at stop}1\} = \Pr\{\xi_3 = 0, \xi_2 = 1 \mid \xi_1 = 1\}$$

$$= \Pr\{\xi_3 = 0 \mid \xi_1 = 1, \xi_2 = 1\} \Pr\{\xi_2 = 1 \mid \xi_1 = 1\} = q_3(1 - q_2) .$$

In general, at stop *i* (*i*=1,…,*N*-1), we have

$$p_{i(i+1)} = q_{i+1} \qquad \text{and} \qquad p_{ij} = q_j \prod_{k=i+1}^{j-1} (1 - q_k) \qquad (j=i+2,\ldots,N). \tag{4}$$

Once the alighting probability matrix has been calculated, we may reconstruct the actual counts of passengers boarding at *i* and alighting at *j* as follows:

$$\hat{x}_{ij} = p_{ij} y_i \qquad (i=1,\ldots,N\text{-}1;\ j=i+1,\ldots,N), \tag{5}$$

with $\hat{x}_{ij} = 0$ if $i \geq j$.

It should be noted that the above Markov model assumes that once a passenger has boarded the transit vehicle, he/she 'forget' his/her origin in terms of choice of destination. From a practical point of view, this assumption may not be realistic and thus too restrictive for some applications. In particular, for those passengers who stay on the transit vehicle for exactly one stop, the alighting probabilities may be unrealistic under the Markov model if the two stops are very close. Extensions will be considered in Section 5.

3. Bayesian Inference

In this section we carry out Bayesian analysis to estimate $\mathbf{P} = [p_{ij}]$ using the prior information and collected data, i.e. on-off counts of passengers, y_i and z_j ($i,j=1,\ldots,N$).

3.1. Likelihood

We first note that immediately before the transit vehicle reaches stop *j* ($j=2,..,N$), the number of passengers on board is $\sum_{k=1}^{j-1}(y_k - z_k)$. Due to the nature of the Markov model, the number of passengers alighting at stop *j*, i.e. z_j, follows a binomial distribution $Bin(\sum_{k=1}^{j-1}(y_k - z_k), q_j)$. The maximum likelihood estimate of q_j is thus given by

$$\hat{q}_j^{ML} = z_j / \sum_{k=1}^{j-1}(y_k - z_k) \qquad j=2,\ldots,N\text{-}1. \tag{6}$$

3.2. Posterior Distribution

To complete the specification for Bayesian analysis, we consider the following conjugate prior for q_j:

$$q_j \sim beta(\alpha_j, \beta_j) \tag{7}$$

where $beta(\alpha, \beta)$ is a beta distribution with a mean of $\alpha/(\alpha+\beta)$ and a variance of $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$. The hyper-parameters α_j and β_j are determined using prior knowledge of q_j. In the case where no prior knowledge is available, we take a non-informative prior, $\alpha_j = 1$ and $\beta_j = 1$, so that the prior of q_j is a uniform distribution on the interval [0, 1].

Applying Bayes' rule to combine the likelihood $Bin(\sum_{k=1}^{j-1}(y_k - z_k), q_j)$ with the prior (7), the posterior distribution of q_j is given by

$$q_j \,|\, (y_k, z_k \text{ for all } k) \sim beta(\alpha_j + z_j, \beta_j + \sum_{k=1}^{j-1}(y_k - z_{k+1})) \qquad j=2,\ldots,N\text{-}1. \tag{8}$$

In Bayesian analysis, the estimate of a parameter is usually taken as its posterior mean. Calculating the mean of the posterior distribution (8) yields an estimate of q_j:

$$\hat{q}_j = (\alpha_j + z_j)/(\alpha_j + \beta_j + \sum_{k=1}^{j-1}(y_k - z_k)) \qquad j=2,\ldots,N\text{-}1. \tag{9}$$

Once the Markov transition probabilities are estimated as equation (9), the alighting probability matrix $\mathbf{P} = [p_{ij}]$ and the trip matrix $\mathbf{X} = [x_{ij}]$ can be calculated by substituting equation (9) into equations (4) and (5) respectively.

We also note that when several (say A) independent samples of on-off counts are drawn, $\{y_1^{(k)}, ..., y_{N-1}^{(k)}, z_2^{(k)}, ..., z_N^{(k)}\}$ (k=1,…,A), the on-off counts y_i and z_j in equation (9) may be replaced by the corresponding aggregated total counts, $\sum_{k=1}^{A} y_i^{(k)}$ and $\sum_{k=1}^{A} z_j^{(k)}$ respectively.

3.3. Posterior Distribution Simulation

The previous analysis focuses on the point estimation of parameters only. However, the joint posterior distribution of the entries of $\mathbf{P} = [p_{ij}]$ or $\mathbf{X} = [x_{ij}]$ can be simulated easily as outlined below. First we draw samples of q_j from equation (8). Then samples of p_{ij} and x_{ij} can be calculated using equation (4) and (5), upon which a joint posterior distribution of p_{ij} and x_{ij} can be calculated.

On the basis of simulated posterior distributions, Bayesian credible intervals for any of the parameters of interest, say the entries of $\mathbf{P} = [p_{ij}]$ or $\mathbf{X} = [x_{ij}]$, can be easily calculated.

4. Relationship with the Maximum Entropy Approach

In this section we explore the relationship between the method developed in the previous sections and the maximum entropy approach.

The maximum entropy approach was developed to reconstruct the actual number of trips x_{ij} for a general trip matrix (Van Zuylen and Willumsen, 1980). Mathematically, the reconstructed numbers of trips T_{ij} are the solution of an optimization problem having a criterion of entropy constrained by observed trip counts. For the problem of transit route matrices, this optimization problem reduces to:

$$\max_{T_{ij} \text{ for } j>i} -\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (T_{ij} \log T_{ij} - T_{ij}), \tag{10}$$

$$\text{subject to } \sum_{j=i+1}^{N} T_{ij} = y_i \qquad (i=1,\ldots,N\text{-}1), \tag{11}$$

$$\sum_{i=1}^{j-1} T_{ij} = z_j \qquad (j=2,\ldots,N). \tag{12}$$

The solution obtained by Van Zuylen and Willumsen (1980) for a general O-D matrix does not apply to transit route matrices because there are some extra constraints for problem

(10)-(12), i.e., $T_{ij} = 0$ for all $i \geq j$. The solution to problem (10)-(12) is given by the following theorem. See the Appendix for the proof.

Theorem 1. The optimal solution to problem (10)-(12) is given by:

$$\hat{T}_{ij}^{ME} = \begin{cases} a_i b_j & \text{if } j > i \\ 0 & \text{otherwise} \end{cases}, \tag{13}$$

where $a_i > 0$ and $b_j > 0$ satisfy the following recursive equations

$$a_i = \xi_i a_{i+1} \quad (i=1,\ldots,N\text{-}2), \tag{14}$$

and $$b_j = \eta_j b_{j-1} \quad (j=N,\ldots,3). \tag{15}$$

The coefficients ξ_i and η_j are given by

$$\xi_1 = \frac{y_1 - z_2}{y_2} \quad \text{and} \quad \xi_i = \frac{y_i}{y_{i+1}} \frac{\sum_{m=1}^{i}(y_m - z_m) - z_{i+1}}{\sum_{m=1}^{i}(y_m - z_m)} \quad (i=2,\ldots,N\text{-}2), \tag{16}$$

$$\eta_N = \frac{z_N - y_{N-1}}{z_{N-1}} \quad \text{and} \quad \eta_j = \frac{z_j}{z_{j-1}} \frac{\sum_{m=1}^{j-2}(y_m - z_m) - z_{j-1}}{\sum_{m=1}^{j-1}(y_m - z_m)} \quad (j=N\text{-}1,\ldots,3). \tag{17}$$

Now let $\hat{x}_{ij}^{ML} = \hat{p}_{ij}^{ML} y_i$ denote the reconstructed number of trips via the maximum likelihood method, where $\hat{p}_{i(i+1)}^{ML} = \hat{q}_{i+1}^{ML}$ and $\hat{p}_{ij}^{ML} = \hat{q}_j^{ML} \prod_{k=i+1}^{j-1}(1 - \hat{q}_k^{ML})$ ($j=i+2,..,N$). $\hat{q}_j^{ML}$ is the maximum likelihood estimate given by equation (6). The relationship between $\hat{x}_{ij}^{ML}$ and $\hat{T}_{ij}^{ME}$ is given in the following theorem. See the Appendix for the proof.

Theorem 2. In the case of no available prior information, the reconstructed number of trips $\hat{T}_{ij}^{ME}$ by the maximum entropy approach is the same as the maximum likelihood estimate $\hat{x}_{ij}^{ML}$:

$$\hat{T}_{ij}^{ME} = \hat{x}_{ij}^{ML} \qquad \text{for all } i \text{ and } j.$$

Theorem 2 has an important practical implication. When no prior information is available, the use of the maximum entropy method for the estimation of transit route O-D matrices is equivalent to the maximum likelihood estimation under a very strong assumption on the structure of the O-D matrices, i.e. the 'forgetfulness property' of the Markov model. Consequently, for the estimation problem of transit route O-D matrices, anyone who has doubts about the practical applicability of the first-order Markov model should have the same concerns about using the maximum entropy method.

5. Extensions

In this section we extend statistical inference for transit route O-D matrices to an *m*th-order Markov model ($2 \le m \le N$). For simplicity, we focus on the case of $m = 2$ but the developed method can be easily generalized to a higher order.

5.1. The Second-Order Markov Model

A first-order Markov model may give undue weight to very short trips. A natural extension is to assume that the probability of alighting at a stop depends on the passenger's status over the past two stops. In doing so, the probability of staying on the transit vehicle for exactly one stop can be modeled separately from the rest of the alighting probabilities. This leads to a second-order Markov model. Following the notation used in Section 2, we define the Markov transition probabilities of the second-order Markov model as follows:

$$\Pr\{\xi_i = k \mid \xi_{i-1} = 1, \xi_{i-2} = 1\} = \begin{cases} q_i & \text{if } k = 0 \\ 1 - q_i & \text{if } k = 1 \end{cases} \qquad (i=3,\ldots,N),$$

where q_i is the probability that a passenger will alight at stop *i* given that he/she is on board at both stops $i-1$ and $i-2$. In addition, we have $q_2 = 0$ by definition.

To address the issue of undue weight for very short trips, we define the conditional probability that a passenger stays on the transit vehicle for exactly one stop:

$$\Pr\{\xi_i = k \mid \xi_{i-1} = 1, \xi_{i-2} = 0\} = \begin{cases} r_i & \text{if } k = 0 \\ 1 - r_i & \text{if } k = 1 \end{cases} \qquad (i=2,\ldots,N),$$

where r_i is the probability that a passenger will alight at stop i given that he/she boarded at stop $i-1$. Clearly we have $q_N = 1 - r_N$.

The alighting probability matrix $\mathbf{P} = [p_{ij}]$ can be calculated as follows:

$$p_{i(i+1)} = r_{i+1} \quad \text{and} \quad p_{ij} = q_j (1 - r_{i+1}) \prod_{k=i+2}^{j-1} (1 - q_k) \qquad (i=1,\ldots,N\text{-}1; j=i+2,\ldots,N). \quad (18)$$

Then the actual counts of passengers boarding at i and alighting at j can be reconstructed:

$$\hat{x}_{ij} = p_{ij} y_i \qquad (i=1,\ldots,N\text{-}1;\ j=i+1,\ldots,N). \quad (19)$$

5.2. 'Complete-Data' Posterior Distribution

Immediately before arriving at stop j (j=2,..,N), the number of passengers on board is $M_j = \sum_{k=1}^{j-1} (y_k - z_k)$. Now let s_j denote the *unobservable* number of the passengers boarding at stop $j-1$ and alighting at stop j (j=3,..,N-1). Hence, $z_j - s_j$ is the number of the passengers who alight at stop j but boarded at stop $j-2$ or earlier. $M_j - z_j$ is the number of passengers who do not alight at stop j. The triplet $(s_j, z_j - s_j, M_j - z_j)$ forms 'complete' data for statistical inference, which follows a trinomial distribution $Mult(M_j, \boldsymbol{\theta}_j)$ with $\boldsymbol{\theta}_j = [r_j, q_j, 1 - r_j - q_j]^T$ due to the nature of the second-order Markov model.

Now we consider a conjugate prior for $\boldsymbol{\theta}_j$, Dirichlet distribution $\boldsymbol{\theta}_j \sim Dirichlet(\alpha_{j1}, \alpha_{j2}, \alpha_{j3})$, where all the hyper-parameters α_{jk} are determined using prior knowledge of $\boldsymbol{\theta}_j$ (j=3,..,N-1) (see, e.g. Schafer (1997, Section 7.2) for the definition of

Dirichlet distributions). When no prior knowledge is available, we take a non-informative prior, $\alpha_{jk} = 1$, so that the priors are uniform. Applying Bayes' rule to combine the likelihoods with the priors, the posterior distributions of $\boldsymbol{\theta}_j$ (j=3,..,N-1) are given by

$$\boldsymbol{\theta}_j \,|\, (s_k, y_k, z_k \text{ for all } k) \sim Dirichlet(\alpha_{j1} + s_j, \alpha_{j2} + (z_j - s_j), \alpha_{j3} + M_j - z_j). \qquad (20)$$

Finally we note that all passengers alighting at stop j=2 stay on the transit vehicle for one stop only. So the number of passengers alighting at stop j=2 follows a binomial distribution $Bin(y_1, r_2)$. On the other hand, all passengers alight at stop N so that the number of the passengers alighting at stop N given that they boarded at stop $N-1$ follows a binomial distribution $Bin(z_N, r_N)$. Then for the conjugate priors chosen as beta distributions, $r_2 \sim beta(\alpha_{21}, \alpha_{22})$ and $r_N \sim beta(\alpha_{N1}, \alpha_{N2})$, the posterior distributions of r_2 and r_N are

$$r_2 \,|\, (y_k, z_k \text{ for all } k) \sim beta(\alpha_{21} + z_2, \alpha_{22} + (y_1 - z_2)), \qquad (21)$$

$$r_N \,|\, (s_k, y_k, z_k \text{ for all } k) \sim beta(\alpha_{N1} + s_N, \alpha_{N2} + (M_N - s_N)). \qquad (22)$$

5.3. Posterior Distribution Simulation

The posterior in (21) is based on observed data only so r_2 can be simulated straightforwardly. The posteriors in (20) and (22), however, are 'complete-data' posterior distributions, i.e. they depend on unobservable 'complete' data s_j (j=3,..,N). For 'complete-data' posteriors, data augmentation (see e.g. Schafer, 1997, Chapter 3) can be used to simulate parameters of interest.

We first note that for given s_j, we may draw $\boldsymbol{\theta}_j$ and r_N from the 'complete-data' posteriors in (20) and (22). On the other hand, for given $\boldsymbol{\theta}_j$ and r_N, the distribution of s_j is a binomial distribution $Bin(z_j, r_j/(r_j + q_j))$ (j=3,…,N) (see e.g. Schafer, 1997, pp 243-244). An

algorithm based on data augmentation for simulating $\boldsymbol{\theta}_j$ and r_N is given below. It simulates the parameters, $\boldsymbol{\theta}_j$ (j=3,..,N-1) and r_N, and the unobservable numbers s_j alternately.

Initialisation: Set initial guess of s_j (j=3,…,N);

Posterior step: draw $\boldsymbol{\theta}_j$ (j=3,..,N-1) from (20) and draw r_N from (22) for given s_j;

Imputation step: draw s_j from $Bin(z_j, r_j/(r_j + q_j))$ (j=3,…,N) for given $\boldsymbol{\theta}_j$ and r_N;

Alternate the Posterior step and Imputation step to convergence.

After the posterior distributions of $\boldsymbol{\theta}_j$, r_2 and r_N have been simulated, the corresponding $\mathbf{P} = [p_{ij}]$ and $\mathbf{X} = [x_{ij}]$ can be calculated using equations (18) and (19). This produces posterior distributions of $\mathbf{P} = [p_{ij}]$, and $\mathbf{X} = [x_{ij}]$ respectively. Upon the posteriors, estimates of the parameters and the corresponding Bayesian credible intervals can be calculated.

5.4. Statistical Inference Based on A Higher-order Markov Model

In general, an *m*th-order Markov model can be used so that the probability of alighting depends on the passenger's status over the past *m* stops. Define complete data consisting of *unobservable* numbers of passengers boarding at stop $j-k$ and alighting at stop j (k=1,…,m; j=3,..,N-1). Then the 'complete-data' likelihood at stop j is an (m+1)-dimensional multinomial distribution. Choosing a prior as a Dirichlet distribution, the posterior also follows a Dirichlet distribution with updated parameters. Similar to Section 5.3, data augmentation can be used to simulate the parameters of interest.

6. A Numerical Example

In this section we illustrate the developed method using a practical example. Li and Cassidy (2007) investigated a bus route served by AC Transit. This route is 26 km in length and serves 58 stops in total. On-off counts of passengers for six trips were collected during a three-hour-long morning peak.

This data set was re-analyzed as follows. To illustrate the Bayesian approach, we first carried out a preliminary analysis where the data collected in the first two trips were used to estimate the Markov transition probabilities, the alighting probability matrix, and the trip matrix via a Bayesian analysis with a non-informative prior. The resulting estimates were considered as prior information in the subsequent analysis.

Next, a Bayesian analysis using the data collected from the last four trips was carried out. Figure 2 displays prior estimates (taken as prior means) of the Markov transition probabilities q_i (i=2,…,57) and the envelop of associated 95% credible intervals (the dotted lines) in the Bayesian analysis. In addition, Figure 3 displays the corresponding posterior estimates (taken as posterior means) and the envelop of associated 95% credible intervals (the dotted lines). Comparing Figure 3 with Figure 2, it can be seen that the posterior credible intervals are much narrower, indicating that the quality of the estimates was greatly improved once the data collected from the last four trips were used. This can also be seen from Figure 2 where the Markov transition probabilities for the last ten or so bus stops were considerably overestimated *a priori*. They were then adjusted to some extent when more information became available from the last four trips, as indicated by the posterior estimates in Figure 3.

(Figure 2 is about here)

Figure 2. The prior estimates of the Markov transition probabilities (real line) and the envelop of associated 95% credible intervals (dotted line)

(Figure 3 is about here)

Figure 3. The posterior estimates of the Markov transition probabilities (real line) and the envelop of associated 95% credible intervals (dotted line)

7. Discussion and Conclusions

For on-off counts of passengers, a parametric approach to statistical inference for transit route O-D matrices has been investigated via the first-order Markov model. A Bayesian method was incorporated to draw statistical inference about the parameters of the first-order Markov model by combining prior information with current observations, leading to a closed-form solution given by equation (9). This method has a close link with the maximum entropy approach: when prior information is not available, the two approaches produce the same estimate of a transit route trip matrix. This indicates that when no prior information is available, there is a close link between the maximum entropy method and the implicit assumption on the 'forgetfulness' in the first-order Markov model.

In many applications prior information about a transit route matrix is available, so the efficiency of utilizing prior information is an important issue. In contrast to conventional approaches such as the balancing method (Lamond and Stewart, 1981), the method developed in this paper incorporates Bayesian analysis, an efficient way to make use of prior information. Consequently the problem of non-structural zeros is completely circumvented. It also allows a full Bayesian analysis of the quantities of interest, such as Markov transition probabilities, alighting probability matrices, and trip matrices, via the obtained joint posterior distributions.

From a computational perspective, the developed method admits a closed-form estimate of a transit route matrix that can be calculated non-iteratively. It is thus computationally more efficient than iterative methods such as the balancing method (Lamond and Stewart, 1981), the MCMC approach (Hazelton, 2001), and the EM algorithm (Li, 2005).

When the first-order Markov model is too restrictive for some applications, we have shown that an *m*th-order Markov model may be used, where the probability of alighting depends on the passenger's status over the past *m* stops. It is worth noting that an *N*th-order Markov model virtually becomes to the general scenario where no particular structure is imposed on O-D matrices. Hence, with a relatively high order *m*, an *m*th-order Markov model can approximate the reality reasonably well. The price to pay, however, is that the total number of the *m*th-order Markov transition probabilities will increase rapidly when the order *m* becomes large.

One of the referees suggests that the passenger's status may be modeled by some additional variables such as passengers' trip purposes. Undoubtedly using additional information like trip purposes will greatly improve the quality of the statistical inference about transit route OD matrices. For this purpose, a much more comprehensive survey has to be carried out at each stop, which leads to a different assumption about the data structure from that assumed in this paper, i.e. on-off counts. This approach will be explored in our future research.

Acknowledgements

The author would like to thank Dr Yuwei Li and Professor Michael J. Cassidy in the Department of Civil and Environmental Engineering at University of California, Berkeley, for providing the AC Transit data. The author gratefully acknowledges the referees' constructive comments and their valuable suggestions for improving earlier versions of the paper.

Appendix. Proofs of Theorems

Proof of Theorem 1. Define the Lagrangian for problem (10)-(12) as

$$L = -\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (T_{ij} \log T_{ij} - T_{ij}) - \sum_{i=1}^{N-1} \lambda_i \{\sum_{j=i+1}^{N} T_{ij} - y_i\} - \sum_{j=2}^{N} \mu_j \{\sum_{i=1}^{j-1} T_{ij} - z_j\}.$$

Differentiating the Lagrangian with respect to T_{ij} and equating them to zero yield the first-order conditions. It is easy to verify that the optimal solution given by $\hat{T}_{ij}^{ME} = a_i b_j$ (for $j > i$) satisfies these first-order conditions, where $a_i = \exp(-\lambda_i)$ and $b_j = \exp(-\mu_j)$. Substituting $\hat{T}_{ij}^{ME}$ into equations (1) and (2) we have

$$a_i \sum_{j=i+1}^{N} b_j = y_i, \qquad i=1,\ldots,N\text{-}1, \tag{A1}$$

and $$b_j \sum_{i=1}^{j-1} a_i = z_j, \qquad j=2,\ldots,N. \tag{A2}$$

Next, we show the recursive equation (14). The proof is by induction. First, we consider the case of i=1. We note that equations (A1) and (A2) with i=1 and j=2 are

$$a_1(b_2 + b_3 + ... + b_N) = y_1, \tag{A3}$$

and $$b_2 a_1 = z_2. \tag{A4}$$

Substituting (A4) into (A3) we obtain

$$a_1(b_3 + ... + b_N) = y_1 - z_1.$$

In addition, equation (A1) with i=2 is $a_2(b_3 + ... + b_N) = y_2$. Combining these two equations we have $a_1 = \{(y_1 - z_2)/y_2\}a_2$. Hence, equation (14) holds for i=1.

Now suppose that equation (14) holds for $i=k$ such that $a_k = \xi_k a_{k+1}$ with $\xi_k = \frac{y_k}{y_{k+1}} \frac{\sum_{m=1}^{k}(y_m - z_m) - z_{k+1}}{\sum_{m=1}^{k}(y_m - z_m)}$. We consider the case of $i=k$+1. Equation (A2) with $j=k$+2 is

$$b_{k+2}(a_1 + ... + a_k + a_{k+1}) = z_{k+2}. \tag{A5}$$

It is easy to verify the following identity:

$$a_1 + \ldots + a_k = (a_k / y_k)\sum_{m=1}^{k}(y_m - z_m).$$

Inserting equation $a_k = \xi_k a_{k+1}$ and the above identity into (A5), we obtain

$$b_{k+2} a_{k+1} = z_{k+2} y_{k+1} / \sum_{m=1}^{k+1}(y_m - z_m). \tag{A6}$$

We note that equation (A1) with $i=k+1$ is

$$a_{k+1}(b_{k+2}+...+b_N)=y_{k+1}. \tag{A7}$$

So substituting (A6) into (A7) we obtain

$$a_{k+1}(b_{k+3}+\ldots+b_N)=y_{k+1}\{\textstyle\sum_{m=1}^{k+1}(y_m-z_m)-z_{k+2}\}/\sum_{m=1}^{k+1}(y_m-z_m). \tag{A8}$$

On the other hand, equation (A1) with $j=k+2$ is

$$a_{k+2}(b_{k+3}+...+b_N)=y_{k+2}. \tag{A9}$$

Combining (A8) and (A9) we obtain $a_{k+1}=\xi_{k+1}a_{k+2}$ with $\xi_{k+1}=\dfrac{y_{k+1}}{y_{k+2}}\dfrac{\sum_{m=1}^{k+1}(y_m-z_m)-z_{k+2}}{\sum_{m=1}^{k+1}(y_m-z_m)}$.

Hence, equation (14) holds for $i=k+1$. By the induction principle, recursive equation (14) holds for any $i=1,\ldots,N-1$.

The proof for equation (15) is similar. This completes the proof.

Remark: Although solution $\hat{T}_{ij}^{ME}$ can be uniquely determined by the first order conditions, a_i and b_j cannot: if $(a_i^*,\ b_j^*)$ is a solution, so does $(\rho a_i^*,\ b_j^*/\rho)$ for any $\rho>0$. In practice, we may simply choose any positive numbers as initial values, say $\tilde{a}_{N-1}=1$ and $\tilde{b}_2=1$, and apply recursive equations (14) and (15) to calculate $\tilde{a}_i$ ($i=1,\ldots,N-2$) and $\tilde{b}_j$ ($j=3,\ldots,N$). Then estimate $\hat{T}_{ij}^{ME}$ is proportional to $\tilde{a}_i\tilde{b}_j$, up to a constant, say ρ. The constant ρ can be determined by substituting $\hat{T}_{ij}^{ME}=\rho\tilde{a}_i\tilde{b}_j$ into equation (A4).

Proof of Theorem 2. It is easy to verify that $\hat{T}_{ij}^{ME}/\hat{T}_{i(j-1)}^{ME}=\hat{q}_j^{ML}(1-\hat{q}_{j-1}^{ML})/\hat{q}_{j-1}^{ML}=\eta_j$, where η_j is given by equation (17). Hence, we have $\hat{T}_{ij}^{ME}=\eta_j\hat{T}_{i(j-1)}^{ME}$, so $\hat{T}_{ij}^{ME}$ satisfies the same recursive equation, equation (15), as b_j. Consequently, to prove $\hat{x}_{ij}^{ML}=a_ib_j=\hat{T}_{ij}^{ME}$ for any fixed i and

$j=i+1,\ldots,N$, we only have to show this is true for the initial value of each row i, i.e. $a_i b_{i+1}$ is equal to $\hat{x}^{ML}_{i(i+1)}$. For this end, we note that we have the following identity:

$$b_{i+1} + \ldots + b_N = (b_{i+1} / z_{i+1}) \sum_{m=1}^{i} (y_m - z_m).$$

Substituting this identity into equation (A1), $a_i \sum_{j=i+1}^{N} b_j = y_i$, we obtain $a_i b_{i+1}$ $= y_i z_{i+1} / \sum_{k=1}^{i} (y_k - z_k) = \hat{x}^{ML}_{i(i+1)}$. This completes the proof.

References

Ben-Akiva, M., Macke, P., Hsu, P., 1985. Alternative methods to estimate route level trip tables and expand on-board surveys. Transportation Research Record 1037, 1-11.

Hazelton, L. M., 2001. Inference for origin-destination matrices: estimation, prediction, and reconstruction. Transportation Research Part B 35 (7), 667-676.

Lamond, B., Stewart, N. F., 1981. Bregman's balancing method. Transportation Research Part B 15 (4), 239-248.

Li, B., 2005. Bayesian inference for origin-destination matrices of transport networks using the EM algorithm. Technometrics 47 (4), 399-408.

Li, B., De Moor, B., 1999. Recursive estimation based on the equality-constrained optimization for intersection origin-destination matrices. Transportation Research Part B 33 (3), 203-214.

Li, B., De Moor, B., 2002. Dynamic identification of origin-destination matrices in the presence of incomplete observations. Transportation Research Part B 36 (1), 37-57.

Li, Y., Cassidy, M. J., 2007. A generalized and efficient algorithm for estimating transit route ODs from passenger counts. Transportation Research Part B 41 (1), 114-125.

Maher, M. J., 1983. Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. Transportation Research Part B 17 (6), 435-447.

Schafer, J. L., 1997. Analysis of Incomplete Multivariate Data. Chapman & Hall, London.

Van Zuylen, H. J., Willumsen, L. G., 1980. The most likely trip matrix estimated from traffic counts. Transportation Research Part B 14 (3), 281-293.

Table 1. Illustration of counts of passengers

origin	destination										row total
	1	2	3	…	i-1	i	i+1	…	N-1	N	
1		x_{12}	x_{13}	…	$x_{1(i-1)}$	x_{1i}	$x_{1(i+1)}$	…	$x_{1(N-1)}$	x_{1N}	y_1
2			x_{23}	…	$x_{2(i-1)}$	x_{2i}	$x_{2(i+1)}$	…	$x_{2(N-1)}$	x_{2N}	y_2
…					…	…	…	…	…	…	…
i-1						$x_{(i-1)i}$	$x_{(i-1)(i+1)}$	…	$x_{(i-1)(N-1)}$	$x_{(i-1)N}$	
i							$x_{i(i+1)}$	…	$x_{i(N-1)}$	x_{iN}	y_i
i+1								…	$x_{(i+1)(N-1)}$	$x_{(i+1)N}$	
…									…	…	…
N-1										$x_{(N-1)N}$	y_{N-1}
N											$y_N \equiv 0$
column total	$z_1 \equiv 0$	z_2	z_3	…	z_{i-1}	z_i	z_{i+1}	…	z_{N-1}	z_N	

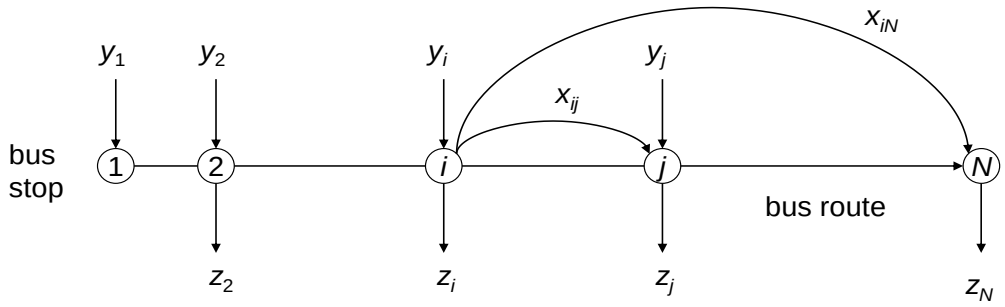

Figure 1. A diagram of a bus transit route serving *N* stops, where x_{ij} are counts of passengers boarding at stop *i* and alighting at stop *j*, y_i and z_i are on and off counts of passengers at stop *i* respectively.

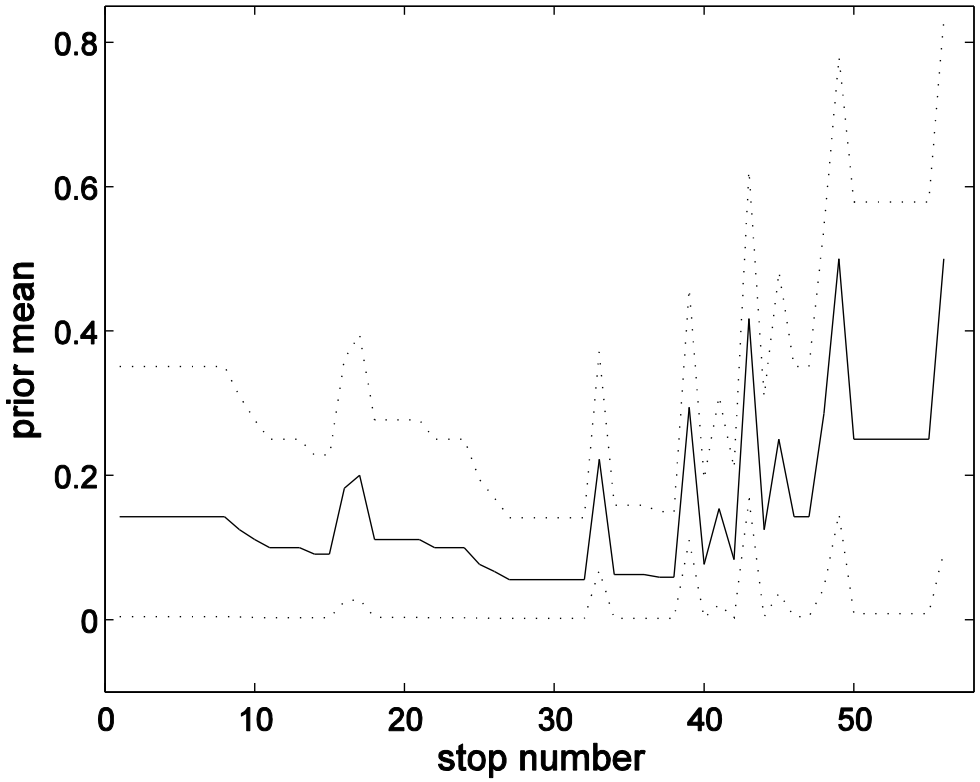

Figure 2. The prior estimates of the Markov transition probabilities (real line) and the envelop of associated 95% credible intervals (dotted line)

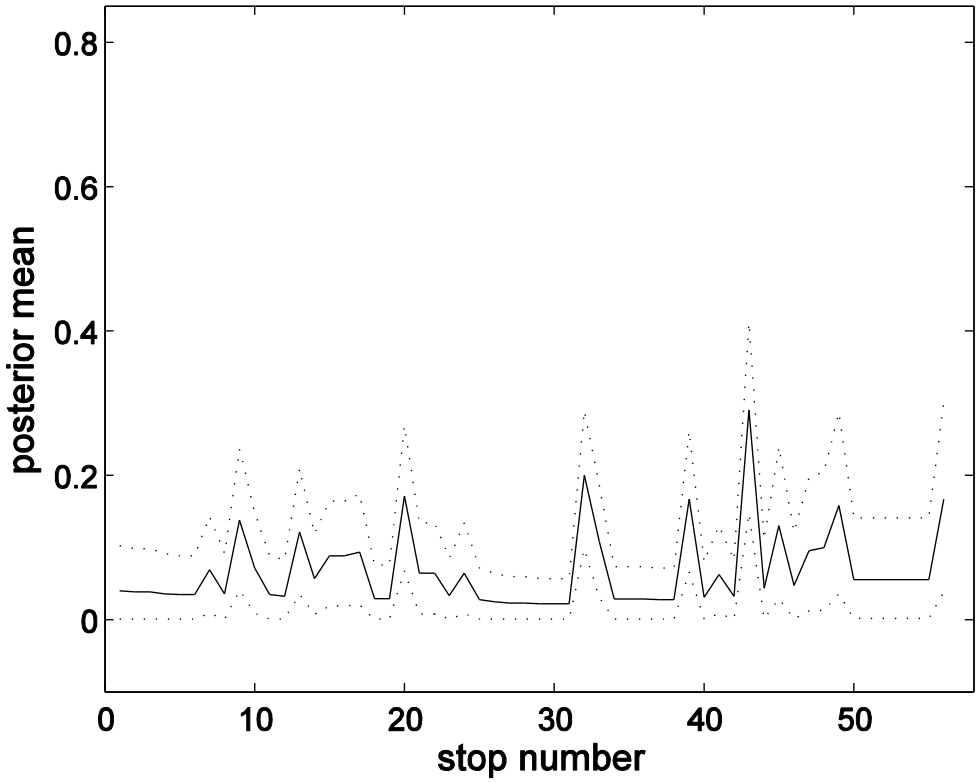

Figure 3. The posterior estimates of the Markov transition probabilities (real line) and the envelop of associated 95% credible intervals (dotted line)