BAYESIAN INFERENCE FOR ORIGIN-DESTINATION MATRICES OF TRANSPORT NETWORKS USING THE EM ALGORITHM

Baibing Li

Business School, Loughborough University
Ashby Road, Loughborough
LE11 3TU, United Kingdom

Email: b.li2@lboro.ac.uk

Summary. Information on the origin-destination (OD) matrix of a transport network is a fundamental requirement in much transportation planning. A relatively inexpensive method to update an OD matrix is to draw inference about the OD matrix based on a single observation of traffic flows on a specific set of network links, where the Bayesian approach is a natural choice to combine the prior knowledge about the OD matrix and the current observation of traffic flows. The existing approaches of Bayesian modeling of OD matrices include using normal approximations to Poisson distributions which leads to the posterior being intractable even under some simple special cases, and/or using MCMC simulation which incurs extreme demand of computational efforts. In this paper, through the EM algorithm, Bayesian inference is reinvestigated for a transport network to estimate the population means of traffic flows, reconstruct traffic flows, and predict future traffic flows. It is shown that the resultant estimates have very simple forms with minimal computational costs incurred.

Keywords: Bayesian statistical modeling; Incomplete data analysis; Multivariate negative binomial distribution; Multivariate Poisson distribution; Trip matrix.

1. Introduction

In this paper we investigate inference about the origin-destination (OD) matrix of a transport network. This is an area that is fundamental in transportation research and has received a lot of attention in the past two or three decades.

Consider a transport network consisting of a number of OD nodes connected through directed links. An OD matrix consists of traffic counts from all origins to all destinations. Historically, trips have been estimated through roadside interviews, number plate surveys, etc. (Watling, 1994; Bierlaire and Toint, 1995), which are expensive in terms of manpower requirements and disruptions of traffic flows. A relatively inexpensive method is to estimate an OD matrix using a single observation of traffic flows on a specific set of network links. The advantages of lower costs and being used for several purposes (accident studies, maintenance planning, etc.) make it very attractive for inference about OD matrices (Van Zuylen and Willumsen, 1980).

There are three major challenges for inference about an OD matrix from a single observation of traffic flows on a specific set of network links. First of all, this is a highly underspecified problem, where the number of links on which measurements of traffic volumes are made is typically much less than the number of unknown parameters of interest. A consequence is that we cannot uniquely determine these unknown parameters based on the collected data solely. In addition, inference based on a single observation excludes the use of asymptotic methods. This can sometimes cause difficulties, for instance, in calculation of posterior variances via the supplemented EM (SEM) algorithm (McLachlan, 1996). Secondly, under the commonly-used assumptions in transport research, traffic volumes measured on the monitored network links have multivariate Poisson distributions (likelihood) and multivariate negative binomial distributions (marginal distributions). These multivariate distributions are very complicated and analytically intractable (Johnson et al. 1997). Finally, dimensions of transport networks are extremely high in most of applications. Computational cost is thus always an issue for the researches in this area.

There is an extensive literature regarding inference about OD matrices within the transport research area. Zuylen and Willumsen (1980) developed an entropy maximizing method to deal with the highly underspecified problem. Maher (1983) considered using a Bayesian approach to combine prior information on an OD matrix with current observations of traffic flows on monitored links. Cascetta (1984) presented a generalized least squares estimator of an OD matrix. In addition, a number of dynamic estimation methods for a special type of OD matrices, intersection OD matrices, have been developed, including Cremer and Keller (1987), Nihan and Davis (1987), Bell (1991), Sheralli et al. (1997), Li and De Moor (1999, 2002), where the major issue is to develop fast algorithms to estimate intersection OD matrices for on-line signal control.

Lo et al. (1996) are the pioneer researchers who recognize the distinction between the estimation of population parameters and the reconstruction of traffic flows. More recently, Hazelton (2001b) investigated some fundamental issues and clarified some confusion in the inference for OD matrices. He clearly defined the following concepts:

Reconstruction: the aim is to estimate the actual number of trips between each OD pair that occurred during the observational period.

Estimation: the aim is to estimate the expected number of OD trips.

Prediction: the aim is to estimate future OD traffic flows.

According to the above definitions, most of the previous researches, except for those of the dynamic estimation methods of intersection OD matrices, are about the reconstruction of traffic flows. Somewhat surprisingly, Hazelton (2001b) shows that the dissimilarity between solutions to a reconstruction problem and an estimation problem is potentially unbounded. Hazelton’s milestone work (2001b) has thus set up a paradigm for investigation of OD matrices of transport networks.

On the other hand, considerable attention has recently been paid to the estimation of OD matrices in the statistical literature. Vardi (1996) investigated maximum likelihood estimation and considered using the EM algorithm but not in the Bayesian framework. Noting the difficulties of calculating conditional expectations in the EM algorithm, he presented an estimation approach based on the method of moments. Tebaldi and West (1998) investigated a Bayesian inference using MCMC simulation that combines Metropolis-Hastings steps within an overall Gibbs sampling framework. Hazelton (2001a) investigated an application to a particular region of Leicester via a Bayesian inference using MCMC. He also noted the computational difficulties caused by normal approximations (2001 a, b).

Many of the researches in the statistical literature discussed the problem in terms of "packets" or "messages" transmitted over a communication network. The basic topological structure of transport networks is similar to that of communication networks. However, due to the rapid development of information technology, multimedia digital signals transmitted over a communication network exhibit some important characteristics which are different from that of traffic flows over a transport network. First of all, statistical models of traffic flows over the two different types of networks are different. Traditionally the statistical model of Poisson distribution is used for traffic flows over a transport network. It has been recognized in the recent decade, however, traffic flows over a multimedia communication network have statistical models other than a Poisson model. Packets transmitted over a multimedia communication network are often highly correlated and have long-range dependence (Paxson and Floyd, 1995; Li and Hwang, 1997). Secondly, due to the reason of cost, updating an OD matrix of a transport network is typically based on a single observation, whilst inference for traffic flows over a multimedia communication network is often based on much more number of observations. The difference in sample size may lead to utility of different statistical methods, for instance, asymptotic methods. Thirdly, the controllability of traffic flows is also different, which may have significant impact on path choice. Over a multimedia communication network traffic flows may be controlled by routers, bridges, etc., whilst traffic flows over transport networks are almost determined by drivers themselves except for some relatively weak restrictions such as speed limits.

This paper is concerned with Bayesian inference about an OD matrix. The existing Bayesian modeling of OD matrices developed by Hazelton (2001b) is based on the measured traffic volumes on monitored links which have intractable multivariate Poisson distributions. To deal with this problem, Hazelton (2001b) considered using multivariate normal approximations to Poisson distributions but the derived normal approximations were still cumbersome for inference. In this paper, instead of drawing inference using traffic counts on monitored links, we first base our investigation on unobservable 'complete' data, and then employ the EM algorithm to obtain an estimate of the OD matrix.

This paper is structured as follows. In section 2 statistical models for inference about an OD matrix are introduced. A Bayesian inference using the EM algorithm is investigated in section 3. In section 4, we calculate incomplete-data posterior variances and derive approximate incomplete-data marginal posterior distributions. Reconstruction and prediction of traffic flows are discussed in section 5. An example is examined to illustrate the developed method in section 6. Finally, major contributions are summarized in section 7.

2. Statistical models

Consider a transport network consisting of nodes and directed links which connect nodes. Nodes in a transport network may be classified into two categories, origin-destination nodes and internal nodes, where origins (destinations) are defined to be the nodes from (to) which traffic flows start (travel), and internal nodes are the remaining nodes which are not origins/destinations. Origins and destinations may correspond to zones, cities, counties, etc. depending on the level of aggregation. Internal nodes are not of direct interest but they play an important role in defining paths of a network, where a path from one origin to a destination is defined to be a sequence of nodes connected in one direction by links.

As an example, we consider a particular region RA in Leicester investigated by Hazelton (2001a). There are six major OD nodes, i.e. nodes 1, 5, 6, 10, 11, and 14, which are associated with major roads and thus of particular interest. Figure 1 shows a simplified form of the road network in region RA where only the major OD nodes are retained and renumbered in comparison with Hazelton's original network. An example of path is the route from origin 1 to destination 5 via nodes 2, 3 and 4 through London road (A6).

INSERT FIGURE 1 HERE

Figure 1. Abstraction of the road network in region RA of Leicester

In this paper the available data for drawing inference about the OD matrix of a transport network is assumed to be a single observation of traffic flows which is collected over a given period of time on some pre-selected network links. Suppose that traffic counts are available on a set of m network links. Following Hazelton's notations (2001b), let $\mathbf{x} = [x_1,..., x_m]^T$ denote the single observation on the m monitored links over a given period. In addition, let $\mathbf{y} = [y_1,..., y_c]^T$ be the unobservable vector of traffic counts on all feasible paths, and $\mathbf{z} = [z_1,..., z_n]^T$ be the vector of OD counts. Note that in practice vector **x** is observable, whilst both vectors **y** and **z** are unobservable.

We consider two statistical models which are commonly used in the transport literature. In the first statistical model, model I, we define an $m \times c$ path-link incidence matrix, **A**, for the monitored links only, whose (i, j)th element is given by

$$a_{ij} = \begin{cases} 1 & \text{link } i \text{ forms part of path } j \\ 0 & \text{otherwise} \end{cases},$$

and define an $n \times c$ matrix, **B**, whose (i, j)th element is given by

$$b_{ij} = \begin{cases} 1 & \text{path } j \text{ connects O-D pair } i \\ 0 & \text{otherwise} \end{cases},$$

where typically m is much less than n, and n is much less than c. Without loss of generality, it is assumed that matrix **A** has a full row rank since redundant rows of **A** have been removed in the stage of design during which a subset of the network links is selected for monitoring.

In general, we have the following two conservation equations of traffic flows:

$$\mathbf{x} = \mathbf{A}\mathbf{y} \ , \tag{1}$$

$$\mathbf{z} = \mathbf{B}\mathbf{y} \ . \tag{2}$$

Under the commonly-used assumption in transport research (Lo et al. 1996; Hazelton, 2001), $y_1,\ldots,y_c$ are independent Poisson random variables with means $\theta_1,...,\theta_c$ respectively. Then according to equation (1), $\mathbf{x}$ has a multivariate Poisson distribution with a mean of $\mathbf{A\theta}$, where $\boldsymbol{\theta} = [\theta_1,...,\theta_c]^T$. Multivariate Poisson distributions have very cumbersome forms of probability mass functions and are hard to deal with analytically. For a comprehensive review on multivariate Poisson distributions, see Johnson et al. (1997).

An alternative statistical model, model II, is based on a proportional assignment matrix $\mathbf{P} = [p_{ij}]$, where p_{ij} is the probability of using link *j* which connects OD pair *i*, and is assumed to be available. The fundamental equations for model II are given by:

$$\mathbf{x} = \mathbf{P}\mathbf{z} \, . \tag{3}$$

A commonly-used assumption is that the elements of OD counts $z_1,..., z_n$ are independent Poisson variates with menas $\mu_1,...,\mu_n$ respectively. Let $\boldsymbol{\mu} = [\mu_1,...,\mu_n]^T$.

Models I and II are closely related (Hazelton, 2001 a). Throughout this paper, our investigation is based on model I. The model II will be discussed very briefly in section 5.3.

Finally we note that inference for $\mathbf{y}$ and $\boldsymbol{\theta}$ is more fundamental than for $\mathbf{z}$ and $\boldsymbol{\mu}$ (Hazelton, 2001 b). This is because from a practical point of view, the former pair define not only mean numbers of OD trips, but also the assignment matrix of path choice probabilities $\mathbf{P} = [p_{ij}]$. Sometimes these probabilities are even of direct interest in themselves. In addition, from a statistical point of view, Bayesian statistical inference for $\mathbf{z}$ and $\boldsymbol{\mu}$ is relatively straightforward based on the inference for $\mathbf{y}$ and $\boldsymbol{\theta}$. In the rest of the paper we concentrate on the inference for $\mathbf{y}$ and $\boldsymbol{\theta}$.

3. Bayesian inference using the EM algorithm

One of the major problems of drawing inference about an OD matrix from a single observation of traffic flows is that typically it is a highly underspecified problem. A Bayesian analysis thus provides a nice research framework by specifying prior which amounts to introducing extra information based on accumulated knowledge. The major difficulty of performing a Bayesian analysis here, as noted by Hazelton (2001b), is that of the very complicated likelihood, multivariate Poisson distribution of the collected data **x**, which is analytically intractable. To overcome this problem, instead of using the observed traffic flows **x**, we first base our investigation on unobservable traffic counts **y** which have very simple likelihood, Poisson distributions, and then obtain estimates of unknown parameters through the EM algorithm. This can largely simplify the Bayesian inference. Using the terminologies of the EM algorithm, vector **x** is termed incomplete data whilst vector **y** complete data.

3.1. Complete-data Bayesian inference

Consider the non-trivial circumstance where matrix **A** does not have any zero-column. Under the model I, traffic counts on all feasible paths are independent Poisson random variables with means $\theta_1,...,\theta_c$ respectively. The complete-data likelihood of $y_1,\ldots,y_c$, is

$$g(\mathbf{y};\boldsymbol{\theta}) = \prod_{j=1}^{c} P(y_j;\theta_j)\,, \qquad (4)$$

where $P(y_j;\theta_j) = \theta_j^{y_j}\exp(-\theta_j)/y_j$ are the Poisson probability mass functions with parameters θ_j. We further assume that $\theta_1,...,\theta_c$ are independent a priori, each of them having a natural conjugate prior distribution, gamma distribution with parameters α_j and β_j:

$$\pi(\theta_j;\alpha_j,\beta_j) = \{\beta_j^{\alpha_j}/\Gamma(\alpha_j)\}\theta_j^{\alpha_j-1}\exp(-\beta_j\theta_j)\,, \qquad (5)$$

where $\Gamma(s)$ represents the gamma function. The corresponding marginal distributions of y_j, termed prior predictive distributions, thus are negative binomial distributions $NB(\alpha_j, \beta_j)$ (Gelman et al., 1995, p49). Denote densities of $NB(\alpha_j, \beta_j)$ as $h(y_j; \alpha_j, \beta_j)$.

Combining the prior with the likelihood, (4) and (5), gives the posterior density of $\boldsymbol{\theta}$ which is a product of some gamma distributions:

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \prod_{j=1}^{c} \pi(\theta_j; \alpha_j, \beta_j) \prod_{j=1}^{c} \theta_j^{y_j} \exp(-\theta_j) / y_j = \prod_{j=1}^{c} \pi(\theta_j; y_j + \alpha_j, 1 + \beta_j) .$$

Hence, the complete-data posterior distributions of θ_j, $p(\theta_j \mid y_j)$, are gamma distributions $\pi(\theta_j; y_j + \alpha_j, 1 + \beta_j)$. The a posteriori most probable estimates of θ_j, $\hat{\theta}_j$, satisfy $d\pi(\theta_j; y_j + \alpha_j, 1 + \beta_j) / d\theta_j = 0$, which yield immediately

$$\hat{\theta}_j = (y_j + \alpha_j - 1) / (1 + \beta_j) . \tag{6}$$

3.2. The EM algorithm

By application of the EM algorithm for the observed incomplete data **x**, the M-step results in the same equation as equation (6), whilst the E-step involves the calculation of conditional expectations, $E_{\boldsymbol{\theta}^{(k)}}\{y_j \mid \mathbf{x}\}$ for j=1,…,c (McLachlan, 1996), where $\boldsymbol{\theta}^{(k)} = [\theta_1^{(k)}, ..., \theta_c^{(k)}]^T$ denotes the kth iteration of $\boldsymbol{\theta}$. Hence, by the EM algorithm, the (k+1)th iteration is given by

$$\hat{\theta}_j^{(k+1)} = (E_{\boldsymbol{\theta}^{(k)}}\{y_j \mid \mathbf{x}\} + \alpha_j - 1) / (1 + \beta_j) . \tag{7}$$

Calculation of the conditional expectations $E_{\boldsymbol{\theta}^{(k)}}\{y_j \mid \mathbf{x}\}$ will be addressed in next sub-section. After convergence, estimates of θ_j can be rewritten as

$$\hat{\theta}_j = (a_j - 1) / b_j , \tag{8}$$

with $a_j = E_{\boldsymbol{\theta}}\{y_j \mid \mathbf{x}\} + \alpha_j$ and $b_j = 1 + \beta_j$.

For the trivial case where the jth column $\mathbf{A}_j$ of $\mathbf{A}$ is a zero-column, the observed traffic counts do not provide any information to update the corresponding estimate of θ_j. It thus remains unchanged and the equation (8) still holds with $a_j = \alpha_j$ and $b_j = \beta_j$.

3.3. The conditional expectations in the E-step

Iterations of the EM algorithm depend on calculation of the conditional expectations $E_{\boldsymbol{\theta}}\{y_j \mid \mathbf{x}\}$ which are extremely hard to calculate as noted by Vardi (1996).

Lemma 1. Suppose that y_j are independent Poisson random variables with means θ_j (j=1,…,c) and $\mathbf{A} = [\mathbf{A}_1,...,\mathbf{A}_c]$ is an $m\times c$ matrix with $\mathbf{A}_j$ the jth column. Let $\mathbf{y} = [y_1,..., y_c]^T$ and $\boldsymbol{\theta} = [\theta_1,...,\theta_c]^T$. Then for a given $m\times 1$ vector $\mathbf{x}$, we have

$$E_{\boldsymbol{\theta}}[y_j \mid \mathbf{Ay} = \mathbf{x}] = \theta_j \frac{\Pr(\mathbf{Ay} = \mathbf{x} - \mathbf{A}_j)}{\Pr(\mathbf{Ay} = \mathbf{x})}.$$

Lemma 2 (Vardi, 1996). Under the assumptions of Lemma 1 and for large θ_j, a normal approximation to the conditional expectation is given by

$$E_{\boldsymbol{\theta}}[y_j \mid \mathbf{Ay} = \mathbf{x}] \approx \theta_j + \mathbf{V}_{yx}\mathbf{V}_{xx}^{-1}(\mathbf{x} - \mathbf{A}\boldsymbol{\theta})$$

where $\mathbf{V}_{yx} = \theta_j \mathbf{A}_j^T$, $\mathbf{V}_{yx} = \mathbf{V}_{xy}^T$, and $\mathbf{V}_{xx} = \mathbf{A}\mathit{diag}\{\theta_1,...,\theta_c\}\mathbf{A}^T$.

The proof of Lemma 1 and further discussion are given in the Appendix. According to Lemma 1, the conditional expectation of y_j given $\mathbf{Ay} = \mathbf{x}$ is equal to the unconditional expectation of y_j, θ_j, multiplied by a ratio of two probabilities. The major advantage of this approach is that it guarantees the resulting conditional expectations of Poisson random variables being non-negative. In the case where traffic counts are large enough, a normal approximation to the joint distribution of y_j and $\mathbf{x}$ may be applied, resulting in Lemma 2. One problem of applying Lemma 2 is that the resultant conditional expectations may be negative values, an issue of concern for by Vardi (1996).

4. The adjusted marginal posterior distributions

The EM algorithm does not provide the variances of estimates. When the number of collected data is relatively small, asymptotic methods for estimating the variances, such as the supplemented EM (SEM) algorithm, are not applicable (McLachlan, 1996). In this section we first investigate how to calculate posterior variances and then discuss how to adjust the complete-data posterior distributions, $p(\theta_j \mid y_j) = \pi(\theta_j; y_j + \alpha_j, 1+\beta_j)$, to derive some approximate distributions to incomplete-data marginal posterior distributions $p(\theta_j \mid \mathbf{x})$.

4.1. Incomplete-data posterior variances

To calculate the incomplete-data posterior variances, we apply the following conditional variance formula (see, for example, Gelman et al., 1995, p20):

$$\mathrm{var}(\theta_j \mid \mathbf{x}) = E[\mathrm{var}(\theta_j \mid y_j) \mid \mathbf{x}] + \mathrm{var}[E(\theta_j \mid y_j) \mid \mathbf{x}] \tag{9}$$

Since the complete-data posterior distributions of θ_j are gamma distributions $\pi(\theta_j; \alpha_j + y_j, 1+\beta_j)$, we have

$$E(\theta_j \mid y_j) = (y_j + \alpha_j)/(1+\beta_j) \quad \text{and} \quad \mathrm{var}(\theta_j \mid y_j) = (y_j + \alpha_j)/(1+\beta_j)^2,$$

which yield

$$\mathrm{var}(\theta_j \mid \mathbf{x}) = \{E(y_j \mid \mathbf{x}) + \alpha_j + \mathrm{var}(y_j \mid \mathbf{x})\}/\{1+\beta_j\}^2, \tag{10}$$

where y_j are independent negative binomial random variables with distributions $NB(\alpha_j, \beta_j)$, and vector $\mathbf{x} = \mathbf{Ay}$ thus has a multivariate negative binomial distribution.

Denote $\mathbf{e}_j$ as a vector having the *j*th entry of one and zeros elsewhere with a suitable dimension. Let $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_c]^T$ and $\boldsymbol{\beta} = [\beta_1, ..., \beta_c]^T$. Let $\nu_j = \alpha_j / \beta_j$ and $\tau_j^2 = \alpha_j(1+\beta_j)/\beta_j^2$ be the expectations and variances of distributions $NB(\alpha_j, \beta_j)$. The main results are summarized below, where $E[y_j \mid \mathbf{x}]$, $\mathrm{var}[y_j \mid \mathbf{x}]$ and $\Pr(\mathbf{Ay} = \mathbf{x})$ which are dependent on the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are written explicitly as $E[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}]$,

$\mathrm{var}[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}]$ and $\Pr(\mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ respectively. See the Appendix for proof of Lemmas 3 and 4.

Lemma 3. Suppose that y_j are independent negative binomial random variables with parameters $\{\alpha_j, \beta_j\}$ (j=1,…,c) and $\mathbf{A} = [\mathbf{A}_1, ..., \mathbf{A}_c]$ is an m×c matrix with $\mathbf{A}_j$ the jth column. Let $\mathbf{y} = [y_1, ..., y_c]^T$. Then for a given m×1 vector, $\mathbf{x}$, we have

(i) $E[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}] = \nu_j \dfrac{\Pr(\mathbf{Ay} = \mathbf{x} - \mathbf{A}_j; \boldsymbol{\alpha} + \mathbf{e}_j, \boldsymbol{\beta})}{\Pr(\mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta})}$;

(ii) $\mathrm{var}[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}] = E[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}]\ \{1 + E[y_j \mid \mathbf{Ay} = \mathbf{x} - \mathbf{A}_j; \boldsymbol{\alpha} + \mathbf{e}_j, \boldsymbol{\beta}]$

$$- E[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}]\}.$$

Lemma 4. Under the assumptions of Lemma 3 and for large α_j, a normal approximation to the conditional expectation and conditional variance is given by

$$E[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}] \approx \nu_j + \mathbf{U}_{yx}\mathbf{U}_{xx}^{-1}(\mathbf{x} - \mathbf{A}\boldsymbol{\nu}),$$

$$\mathrm{var}[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}] \approx \mathbf{U}_{yy} - \mathbf{U}_{yx}\mathbf{U}_{xx}^{-1}\mathbf{U}_{xy},$$

where $\mathbf{U}_{yx} = \tau_j^2 \mathbf{A}_j^T$, $\mathbf{U}_{yx} = \mathbf{U}_{xy}^T$, $\mathbf{U}_{xx} = \mathbf{A}\boldsymbol{\Sigma}\,\mathbf{A}^T$ and $\boldsymbol{\Sigma} = diag\{\tau_1^2, ..., \tau_c^2\}$.

4.2. Incomplete-data marginal posterior distributions

From the results of previous sections, the complete-data posterior distributions, $p(\theta_j \mid y_j)$, are gamma distributions $\pi(\theta_j; y_j + \alpha_j, 1 + \beta_j)$. In practice, for the given incomplete data $\mathbf{x}$, if we use $\pi(\theta_j; y_j + \alpha_j, 1 + \beta_j)$ as the resultant posterior distributions with y_j being replaced by their corresponding conditional expectations $E\{y_j \mid \mathbf{x}\}$, this will typically exaggerate the precision about θ_j due to equation (9). To draw inference such as construction of Bayesian credible intervals, it is necessary to adjust these complete-data posterior distributions. For this end, we consider using $\pi(\theta_j; r_j(E(y_j \mid \mathbf{x}) + \alpha_j), r_j(1 + \beta_j))$ as approximate distributions to incomplete-data marginal posteriors $p(\theta_j \mid \mathbf{x})$, where r_j are positive scalars which are determined such that the approximate distributions $\pi(\theta_j; r_j(E(y_j \mid \mathbf{x}) + \alpha_j), r_j(1+\beta_j))$ have the same variances as $\text{var}(\theta_j \mid \mathbf{x})$, i.e. the re-scaling parameters, r_j, satisfy

$$\{E(y_j \mid \mathbf{x}) + \alpha_j\} / \{r_j(1+\beta_j)^2\} = \{E(y_j \mid x) + \alpha_j + \text{var}(y_j \mid \mathbf{x})\} / \{1+\beta_j\}^2,$$

which yields $r_j = \{E(y_j \mid \mathbf{x}) + \alpha_j\} / \{E(y_j \mid \mathbf{x}) + \alpha_j + \text{var}(y_j \mid \mathbf{x})\}$. The re-scaling thus results in inflated variances since $r_j \le 1$.

The approximate marginal distributions $\pi(\theta_j; r_j(E(y_j \mid \mathbf{x}) + \alpha_j), r_j(1+\beta_j))$ are quite close to the true incomplete-data marginal posterior distributions $p(\theta_j \mid \mathbf{x})$ which are analytically intractable. They have equal mean and variance. As α_j become large, the coefficients of skewness for both distributions are asymptotically equivalent to $2\alpha_j^{-1/2}$. In addition, both distributions approach to the same normal distribution as $\alpha_j \to +\infty$.

One major advantage of the approximate marginal posterior distributions is that they retain the form of conjugate distributions to the complete-data likelihood. In addition, due to simplicity, the approximate marginal posterior distributions provide a convenient way for drawing Bayesian inference such as construction of Bayesian incredible intervals without resorting to simulation methods.

5. Prediction and reconstruction of traffic flows

5.1. Bayesian prediction of future traffic flows

Consider the non-trivial case where matrix **A** does not have any zero-column. For future traffic flows, $\tilde{\mathbf{y}}$, the complete-data posterior predictive distribution is given by $g(\tilde{\mathbf{y}} \mid \mathbf{y}) = \int g(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}$. This results in the complete-data posterior predictive distribution $\prod_{j=1}^{c} h(\tilde{y}_j; \tilde{\alpha}_j, \tilde{\beta}_j)$ which is a product of the probability mass functions

$h(\tilde{y}_j;\tilde{\alpha}_j,\tilde{\beta}_j)$ of distributions $NB(\tilde{\alpha}_j,\tilde{\beta}_j)$. Hence, the complete-data marginal posterior predictive distributions are negative binomial distributions $NB(\tilde{\alpha}_j,\tilde{\beta}_j)$ with $\tilde{\alpha}_j = y_j + \alpha_j$ and $\tilde{\beta}_j = 1+\beta_j$. The mode of the marginal posterior predictive distribution is thus at

$$\tilde{y}_j = \left[(\tilde{\alpha}_j - 1)/\tilde{\beta}_j\right] = \left[(y_j + \alpha_j - 1)/(1+\beta_j)\right],$$

where [u] denotes the integer part of u. Given the incomplete data $\mathbf{x}$, the prediction is

$$\tilde{y}_j = \left[(E\{y_j \mid \mathbf{x}\} + \alpha_j - 1)/(1+\beta_j)\right].$$

For the trivial case that $\mathbf{A}_j = \mathbf{0}$ for some j, the corresponding marginal posterior predictive distribution is still a negative binomial distribution with un-updated parameters, α_j and β_j. In this case, the mode of the marginal posterior predictive distribution is at $\tilde{y}_j = \left[(\alpha_j - 1)/\beta_j\right]$. Therefore, the prediction of future traffic flows may be rewritten in a unified form:

$$\tilde{y}_j = \left[(a_j - 1)/b_j\right], \tag{8$'$}$$

where $a_j = E\{y_j \mid \mathbf{x}\} + \alpha_j$ and $b_j = 1+\beta_j$ if $\mathbf{A}_j \neq \mathbf{0}$; $a_j = \alpha_j$ and $b_j = \beta_j$ otherwise. It shares the same form as equation (8) except that traffic counts have to be taken as integers.

Next we consider posterior predictive variances. Similar to equation (9), the incomplete-data predictive variances are given by

$$\mathrm{var}(\tilde{y}_j \mid \mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta}) = E[\mathrm{var}(\tilde{y}_j \mid y_j) \mid \mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta}] + \mathrm{var}[E(\tilde{y}_j \mid y_j) \mid \mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta}].$$

Since $E(\tilde{y}_j \mid y_j) = \tilde{\alpha}_j/\tilde{\beta}_j$ and $\mathrm{var}(\tilde{y}_j \mid y_j) = \tilde{\alpha}_j(\tilde{\beta}_j + 1)/\tilde{\beta}_j^{\,2}$, we obtain

$$\mathrm{var}(\tilde{y}_j \mid \mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta}) = \{[E(y_j \mid \mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta}) + \alpha_j](2+\beta_j) + \mathrm{var}(y_j \mid \mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta})\}/\{1+\beta_j\}^2, \tag{11}$$

where $\mathrm{var}(y_j \mid \mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta})$ and $E(y_j \mid \mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta})$ may be calculated using Lemmas 3 and 4. In comparison with equation (10), it can be seen that posterior predictive variances are larger than posterior estimation variances.

Since the complete-data marginal posterior predictive distributions, $NB(\tilde{\alpha}_j, \tilde{\beta}_j)$, exaggerate the precision of prediction, we rescale the two parameters of $NB(\tilde{\alpha}_j, \tilde{\beta}_j)$ to derive approximate incomplete-data marginal posterior predictive distributions. Letting

$$r_j = \{E(y_j \mid \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}) + \alpha_j\} / \{E(y_j \mid \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}) + \alpha_j + \text{var}(y_j \mid \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta})\} \leq 1,$$

we approximate posterior marginal predictive distributions $p(\tilde{y}_j \mid \mathbf{x})$ by $NB(r_j E(\tilde{\alpha}_j \mid \mathbf{x}), r_j \tilde{\beta}_j)$ which have the same means and variances as the true distributions.

5.2. Bayesian inference for reconstruction of traffic flows

The problem of reconstruction is to estimate the actual number of trips that occurred during the observational period. From section 3.1, the marginal distributions of y_j are $NB(\alpha_j, \beta_j)$ with densities $h(y_j; \alpha_j, \beta_j)$. The a priori most probable estimates of the traffic flows y_j, $\hat{y}_{0j} = \lfloor (\alpha_j - 1)/\beta_j \rfloor$, are obtained by maximizing the marginal distributions.

For given **y**, the distribution of **x** is degenerate, placing probability one on the single point **Ay**, $I(\mathbf{x} = \mathbf{Ay})$. The posterior density of **y** for given **x** is thus given by

$$p(\mathbf{y} \mid \mathbf{x}) \propto I(\mathbf{x} = \mathbf{Ay}) \prod_{j=1}^{c} h(y_j; \alpha_j, \beta_j). \tag{12}$$

For given observation **x**, the reconstructed traffic flows can be calculated as the a posteriori most probable vector of **y**, i.e. the solution to the following maximization problem:

$$\max_{\mathbf{y}} \prod_{j=1}^{c} h(y_j; \alpha_j, \beta_j) \tag{13}$$

subject to **Ay**=**x** (14)

Similar to Zuylen and Willumsen (1980), the objective function (13) may be approximated by the Stirling's formula:

$$J = \sum_{j=1}^{c} [(y_j + \alpha_j)\log(y_j + \alpha_j) - (y_j + 1)\log(y_j + 1) - y_j \log(1 + \beta_j)].$$

Through the Lagrangean equation, $J - \boldsymbol{\lambda}^T(\mathbf{Ay} - \mathbf{x})$, where $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_m]^T$ is a vector of Lagrangean multipliers, a formal solution is obtained

$$\hat{y}_j = (\alpha_j - 1)/t_j - 1 \approx (\alpha_j - 1)/t_j, \tag{15}$$

where $t_j = (1+\beta_j)\exp(\boldsymbol{\lambda}^T \mathbf{A}_j) - 1$. A special case that $\mathbf{A}_j = \mathbf{0}$ for some j results in $\hat{y}_j = \hat{y}_{0j} = (\alpha_j - 1)/\beta_j$. Note that equation (15) shares the same form as equations (8) and (8)′, $\hat{y}_j \approx (a_j - 1)/b_j$, where $a_j = \alpha_j$. $b_j = (1+\beta_j)\exp(\boldsymbol{\lambda}^T \mathbf{A}_j) - 1$ if $\mathbf{A}_j \neq \mathbf{0}$; $b_j = \beta_j$ otherwise.

The solution in equation (15) may be further approximated as $\hat{y}_j \approx \hat{y}_{0j} \exp(-\boldsymbol{\lambda}^T \mathbf{A}_j)$ when β_j are large. Let $h_i = \exp(-\lambda_i)$, the above equation may be rewritten as

$$\hat{y}_j \approx y_{0j} \prod_{i=1}^{m} h_j^{a_{ij}} \tag{16}$$

which is related to the solution of Zuylen and Willumsen’s maximizing entropy method.

In the case that traffic counts are large such that normal approximations to the negative binomial distributions may be applied, the objective function in (13) can be written approximately as

$$\min_{\mathbf{y}} (\mathbf{y} - \mathbf{v})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{v}) \tag{17}$$

A solution to the problem (17) and (14) is immediately given by

$$\mathbf{y} = \mathbf{v} - \boldsymbol{\Sigma}\mathbf{A}^T(\mathbf{A}\boldsymbol{\Sigma}\,\mathbf{A}^T)^{-1}[\mathbf{Av} - \mathbf{x}]$$

The problem (17) and (14) provides a nice link with Cascetta’s generalized least squares method, Maher’s Bayesian method (1983) and Hazelton’s (2001b) Bayesian method.

5.3. Bayesian inference for a given assignment matrix

Finally, we discuss Bayesian inference about traffic flows under the framework of model II very briefly. Bayesian inference about the mean of $\mathbf{z}$, $\boldsymbol{\mu}$, can be drawn using the EM algorithm in a same way as discussed in section 3. Likewise, prediction of future traffic flows can be carried out as section 5.1. Finally, the reconstructed traffic flows can correspondingly be approximated as $\hat{z}_j \approx z_{0j} \prod_{i=1}^{m} h_j^{p_{ij}}$ with $h_i = \exp(-\lambda_i)$ as its counterpart of equation (16). Note that it has exactly the same form as the Zuylen and Willumsen's solution via maximizing entropy method.

6. Example

To illustrate the developed method, we consider a simple transport network, displayed in Figure 2, which was investigated by Hazelton (2001b). It has six nodes with fourteen directed links, four of which are OD nodes, i.e. nodes 1, 3, 4, and 6. The total number of the OD pairs is twelve. Following Hazelton (2001b), we assume that traffic counts are available on m=8 links, i.e. links 1, 2, 5, 6, 7, 8, 11, 12. As we can see later, this selection covers a number of special cases of interest. For simplicity, we consider fixed routing, assuming that the paths connecting nodes 1 and 6, and paths connecting nodes 3 and 4, are through nodes 2 and 5.

INSERT FIGURE 2 HERE

Figure 2. A simplified network topology

The true OD matrix used to simulate traffic flows in simulation and the prior OD matrix for Bayesian analysis are displayed in columns 2 and 3, Table 1, respectively. In each of the simulation experiments, a vector of traffic flows, **y**, is simulated, whose elements are outcomes of independent Poisson variables with the means given by column 2, Table 1. A single observation **x** on the set of the monitoring links {1, 2, 5, 6, 7, 8, 11, 12} is calculated using equation (1) with the simulated values of **y**.

Table 1. Posterior estimates and the associated variances

OD pair	'True' parameter	Prior estimate	Posterior estimate	Prior variance	Complete data	Incomplete data	Scaling factor

					variance	variance	
1→3	783	793	782.74	793.00	391.62	402.80	0.97
1→4	677	593	616.00	593.00	308.25	308.25	1.00
1→6	137	99	104.26	99.00	52.38	63.59	0.82
3→1	429	526	480.44	526.00	240.47	263.72	0.91
3→4	524	440	440.00	440.00	440.00	440.00	1.00
3→6	104	37	73.50	37.00	37.00	37.00	1.00
4→1	225	269	241.00	269.00	120.75	120.75	1.00
4→3	701	542	542.00	542.00	542.00	542.00	1.00
4→6	30	30	31.24	30.00	15.87	26.96	0.59
6→1	104	138	124.56	138.00	62.53	85.84	0.73
6→3	132	69	100.50	69.00	50.50	50.50	1.00
6→4	81	81	79.44	81.00	39.97	63.16	0.63

6.1. Analysis in one simulation

The prior distributions for θ_j are taken as Gamma distributions with parameters α_j being the prior estimates of θ_j (column 3, Table 1) and $\beta_j = 1$.

Table 1 displays the results in one simulation where the simulated observation vector **x** is [884, 548, 111, 133, 191, 144, 214, 640]^T. It can be seen that the posterior estimates do improve the prior estimates except for OD pairs 3→4 and 4→3. Note that for OD pairs 3→4 and 4→3, this is the completely non-informative situation in the sense that the monitored links, {1, 2, 5, 6, 7, 8, 11, 12}, do not form part of the paths associated with the OD pairs 3→4 and 4→3, and thus the collected data, **x**, do not provide information to update the corresponding estimates. It can also be seen that the resultant incomplete-data posterior variances lie between the prior variances and complete-data posterior variances, indicating the necessity of re-scaling. Note that the re-scaling factor is unity for the OD pairs 1→4, 4→1, 3→6 and 6→3 because the collected data **x** are ‘complete’ for inference about the corresponding population means of traffic flows.

Next, the posterior is explored by using the Metropolis-Hastings algorithm. Implementation of the Metropolis-Hastings algorithm is challenging since it is computationally expensive for evaluation of posterior kernel, in particular, evaluation of mass functions of the multivariate Poisson. The exact method suggested by Tebaldi and West (1998) has high computational costs for large means. In this paper we use the method of normal approximations to Poisson distributions as suggested by Hazelton (2001a). It largely reduces the computational costs in comparison with the method of Tebaldi and West (1998). The proposal in the Metropolis-Hastings algorithm is drawn based on a random walk process with the current values of parameters plus an innovation which is generated using a normal distribution with zero mean and a standard deviation of tuning parameter. For the completely non-informative situation we use their prior distributions for simulation, and those that y_j are completely determined by **x** we use their posterior distributions, gamma distributions, for the simulation. It run for 100 000 iterations and the first 3000 of these were discarded as a burn-in period. To have approximately independent draws from the target distribution after an approximate convergence is reached, we use every 10th simulation draw (Gelman et al., 1995), from which an approximate posterior is obtained.

Table 2 displays a comparison between the method developed here and that of the MCMC. It can be seen that in general these two methods give quite similar results.

6.2. Analysis for repeated simulations

Next, simulation is repeated 500 times to have an overall picture about the performance of the developed method. The quality of prior information varies via adjusting the parameters of the prior distributions. Specifically, the prior distributions of θ_j are taken as Gamma distributions with parameters $\alpha_j = (\eta - 1)\theta_j * + \theta_{j0}$ and $\beta_j = \eta$, where $\theta_j *$ and θ_{j0} are given by the column 2 and 3, Table 1, respectively. η is a constant taken as 1, 2, 5, 10, 20, and 50 respectively. Clearly, taking the value of η =1 leads to prior estimates θ_{j0} with larger variances, whilst taking the value of η =50 results in prior estimates being almost equal to the 'true' mean values with very small variances.

The accuracies of estimation, prediction, and reconstruction in terms of average root mean square error (RMSE) over 500 simulations are displayed in Table 3. The RMSE is defined to be $\| \hat{\mathbf{f}} - \mathbf{f} \|_2 / \sqrt{c}$, where $\hat{\mathbf{f}}$ is an estimate of mean traffic flows, reconstructed traffic flows, and predicted future traffic flows respectively, and **f** is the corresponding true values for the problems of estimation, reconstruction, and prediction respectively.

It can be seen from Table 3 that as the quality of prior information is improved, the values of RMSE for estimation decreases. Not surprisingly, the posterior estimates of the mean traffic flows have much smaller values of RMSE than the corresponding predicted future traffic flows. This can also be seen from equations (10) and (11) that posterior predictive variances are larger than posterior estimation variances.

In addition, Table 3 shows that reconstructed traffic flows have lower values of RMSE than predicted future traffic flows. This is because the reconstruction is a problem where the reconstructed traffic flows have to satisfy some constraints of the observed traffic flows, thus using current information to forecast future values have larger variances.

Finally, from a point of view of predictivism, predicting new observational outcomes of a process has been the principal objective. It can be seen from Table 3 that the accuracy of predictions of future traffic flows is improved as the quality of the prior knowledge about the OD matrix becomes better. Overall the average prediction errors are relatively low in comparison with the magnitudes of the traffic flows.

Table 2. Comparison of posterior means and variances of θ_j (j=1,…,12) using the scaled gamma distributions and MCMC simulation

	OD pair	Posterior mean by scaled gamma	Posterior mean by MCMC	Posterior variance by scaled gamma	Posterior variance by MCMC
j=1	1→3	783.23	782.66	402.80	408.93

j=2	1→4	616.50	616.34	308.25	306.51
j=3	1→6	104.75	104.49	63.59	64.30
j=4	3→1	480.94	481.72	263.72	268.01
j=5	3→4	440.00	439.37	440.00	421.22
j=6	3→6	74.00	73.71	37.00	37.74
j=7	4→1	241.50	241.07	120.75	118.37
j=8	4→3	542.00	541.36	542.00	530.42
j=9	4→6	31.74	31.64	26.96	30.58
j=10	6→1	125.06	125.07	85.84	83.23
j=11	6→3	101.00	100.64	50.50	50.51
j=12	6→4	79.94	79.55	63.16	60.65

Table 3. Average RMSE over 500 simulations

	Estimation	Reconstruction	Prediction
η =1	58.12	52.78	60.33
η =2	31.18	27.32	35.69
η =5	13.48	13.94	21.52
η =10	7.02	10.68	19.10
η =20	3.58	10.54	17.86
η =50	1.45	9.99	17.37

7. Conclusions and discussion

In this paper we have investigated drawing inference about the OD matrix of a transport network from a single observation of traffic flows on a specific set of network links. Besides the issue of computational costs, there are two major theoretical challenges: the problem is highly underspecified and the likelihood is analytically intractable.

Through specifying prior to introduce more information, we have performed a Bayesian analysis to deal with the highly underspecified problem. On the other hand, to overcome the problem of the analytically intractable likelihood, the EM algorithm has been employed, which largely simplifies the Bayesian analysis. Further insights and conclusions have thus been obtained and the resulting simple solutions to the reconstruction problem provide nice links with most of the well-known existing work in the area of transport research.

From a point of view of transport research, our major contribution is the development of a simple but general approach for estimation of OD matrices which includes many previous methods as special cases and has minimal computational cost. Moreover, the neat solutions to the problems of estimation, reconstruction, and prediction show that they are very closely related. All of these solutions share approximately the same form of updating equations.

From a statistical point of view, a Bayesian approach using the EM algorithm is investigated to draw inference about OD matrices of transport networks. To implement this, a simple method is proposed to calculate conditional expectations in the EM algorithm which were considered extremely difficult to deal with (Vardi, 1996). In addition, a re-scaling method is developed to derive approximate incomplete-data marginal posterior distributions which retain conjugate forms to the likelihood of complete data.

Finally, from a computational point of view, the developed approximate posteriors are very competitive in comparison with that of using MCMC simulation, especially for very large-scale transport networks in practice.

Acknowledgement

The author thanks Professor Murray Aitkin for his helpful comments on the earlier version of this paper.

References

Bell, M.G.H. (1991). The estimation of OD matrices by constrained generalized least squares. *Transportation Research*, 25B, 13-22.

Bierlaire, M. and Toint, Ph. L. (1995). MEUSE: an origin-destination matrix estimator that exploits structure. *Transportation Research*, 29B, 47-60.

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimators. *Transportation Research*, 18B, 289-299.

Cremer, M. and Keller, H. (1987). A new class of dynamic methods for the identification of origin-destination flows. *Transportation Research*, 21B, 117-132.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.

Hazelton, L. M. (2001a). Estimation of origin-destination trip rates in Leicester. *Applied Statistics*, 50, 423-433.

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

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete multivariate distributions*. Wiley, New York.

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

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

Li, S.-Q. and Hwang, C.-L. (1997). On the convergence of traffic measurement and wueueing analysis (SMAQ) tool. *IEEE/ACM Trans. Networking*, 5, 95-110.

Lo, H. P., Zhang, N., and Lam, W. H. K. (1996). Estimation of an origin-destination matrix with random link choice proportions: a statistical approach. *Transportation Research*, 30B, 309-324.

Maher, M. J. (1983). Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research*, 17B, 435-47.

McLachlan, G.J. (1996). *The EM algorithm and Extensions*. Wiley, New York.

Nihan, N. and Davis, G. (1987). Recursive estimation of origin-destination matrices from input/output counts. *Transportation Research*, 21B, 149-163.

Paxson, V. and Floyd, S. (1995). Wide Area Traffic: the failure of Poisson modeling. *IEEE/ACM Trans. Networking*, 3, 225-244.

Sherali, H.D., Arora, N. and Hobeika, A.G. (1997). Parameter optimization methods for estimating dynamic origin-destination trip-tables. *Transportation Research*, 31B, 141-157.

Tebaldi, C. and West, M. (1998). Bayesian inference on network traffic using link data (with discussion). *J. of the American Statistical Association*, 93, 557-576.

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

Vardi, Y. (1996). Network tomography: estimating source-destination traffic intensities from link data. *J. of the American Statistical Association*, 91, 365-377.

Watling, D. (1994). Maximum likelihood estimation of an origin-destination matrix from a partial registration plate survey. *Transportation Research*, 28B, 289-314.

Appendix. Conditional expectations and conditional variances

To show Lemma 1, we define a set below:

$$S(\mathbf{x}) = \{u = [u_1,...,u_c]^T \mid u_l\,(l = 1,...,c) \text{ are nonnegative integers satisfying } \mathbf{Au} = \mathbf{x}\}$$

Then according to the definition of expectation we have

$$E_{\boldsymbol{\theta}}[y_j \mid \mathbf{Ay} = \mathbf{x}] = \sum_{\substack{u_l=0\ (l=1,...,c) \\ \text{and } \mathbf{u}\in S(\mathbf{x})}}^{\infty} u_j P(u_j;\theta_j) \prod_{\substack{l=0 \\ l\neq j}}^{c} P(u_l;\theta_l) / \Pr\{\mathbf{Ay} = \mathbf{x}\}$$

Noting the recursion formula $u_j P(u_j;\theta_j) = \theta_j P(u_j - 1;\theta_j)$, and letting $v_j = u_j - 1$ and $v_l = u_l$ (l=1,…,c and l≠j), we obtain

$$E_{\boldsymbol{\theta}}[y_j \mid \mathbf{Ay} = \mathbf{x}] = \theta_j \sum_{\substack{v_l=0\ (l=1,...,c) \\ \text{and } \mathbf{v}\in S(\mathbf{x}-\mathbf{A}_j)}}^{\infty} \prod_{l=1}^{c} P(v_l;\theta_l) / \Pr\{\mathbf{Ay} = \mathbf{x}\}$$

$$= \theta_j \frac{\Pr(\mathbf{Ay} = \mathbf{x} - \mathbf{A}_j)}{\Pr(\mathbf{Ay} = \mathbf{x})} \sum_{\substack{v_l=0\ (l=1,...,c) \\ \text{and } \mathbf{v}\in S(\mathbf{x}-\mathbf{A}_j)}}^{\infty} \prod_{l=1}^{c} P(v_l;\theta_l) / \Pr\{\mathbf{Ay} = \mathbf{x} - \mathbf{A}_j\}$$

Hence, it follows by noting that the sum in the above equation is equal to unity. This completes the proof of Lemma 1.

Two special cases for Lemma 1 are:

(i) the collected data is completely non-informative for inference about θ_j. In this case the monitored links do not form part of the jth path, i.e. $\mathbf{A}_j = \mathbf{0}$. Hence y_j is independent of **x**. We thus obtain $\Pr(\mathbf{Ay} = \mathbf{x} - \mathbf{A}_j) / \Pr(\mathbf{Ay} = \mathbf{x}) = 1$, which yields $E_{\boldsymbol{\theta}}[y_j \mid \mathbf{Ay} = \mathbf{x}] = \theta_j$;

(ii) the collected data is ‘complete’ for inference about θ_j. In this case one of the monitored links, say the kth, does not form part of any path except for the path j, i.e. $\mathbf{A}_j = \mathbf{e}_k$, and the kth row of **A** has the jth entry of one and zeros elsewhere. Hence y_j is the same as the kth entry of **x**, x_k, and y_j is independent of the remaining entries of **x**. We then obtain

$\Pr(\mathbf{Ay}=\mathbf{x}-\mathbf{A}_j)/\Pr(\mathbf{Ay}=\mathbf{x})=\Pr(y_j=x_k-1)/\Pr(y_j=x_k)=x_k/\theta_j$, which yields $E_{\boldsymbol{\theta}}[y_j \mid \mathbf{Ay}=\mathbf{x}]=x_k$.

In the general situation where the problem cannot be reduced to the above special cases, if the volumes of traffic flows are so low such that elements of **x** are small, calculation may be carried out using the exact method discussed in Johnson et al. (1997); otherwise if all of the elements of **y** are large enough such that normal approximations to the joint distributions of y_j and **x** can be applied, Lemma 2 may be used to calculate the required conditional expectations; finally, for the case where all of the elements of **x** are large enough but y_j are not necessarily large, we apply a normal approximation to the multivariate Poisson distribution of **x** and use simple numerical integration over a unit hypercube, yielding:

$$\Pr(\mathbf{Ay}=\mathbf{x})\approx\frac{1}{(2\pi)^{m/2}\,|\mathbf{V}_{xx}|^{1/2}}\exp\{-(\mathbf{x}-\mathbf{A}\boldsymbol{\theta})^T\,\mathbf{V}_{xx}^{-1}(\mathbf{x}-\mathbf{A}\boldsymbol{\theta})/2\} \qquad \text{(A1)}$$

Calculation of ratio of two probabilities in Lemma 1 can be reduced when there are some entries of **x** are 'complete' as defined in case (ii). In this case the corresponding rows of **A** may be deleted because of independence.

For the simplest case where an exact probability of multivariate Poisson distribution is available (Johnson et al. 1997, pp124), numerical simulations were carried out to evaluate the method developed above. For the numerical simulations, the unconditional expectations of individual Poisson variates, $\boldsymbol{\theta}$, were taken randomly from a uniform distribution U[50, 150], and entries of **x** were taken as outcomes of normal random variables with both means and variances equal to the corresponding entries of $\mathbf{A}\boldsymbol{\theta}$. Then conditional expectations were calculated by applying Lemma 1 with exact probabilities and approximate probabilities (A1) respectively. The average relative error of conditional expectations between the exact and approximate methods over 100 simulation experiments was about 3%.

Next, we consider Lemmas 3 and 4. The proof of Lemma 3 can be completed in a similar way to the proof of Lemma 1 by noting the recursion formula

$uh(u;\alpha,\beta) = (\alpha/\beta)h(u-1;\alpha+1,\beta)$. The proof of Lemma 4 is immediate from normal approximations. Calculation of the required probabilities in Lemma 3 can be done in a similar way as outlined above. Two special cases are:

(i) the data is completely non-informative for inference about y_j i.e. $\mathbf{A}_j = \mathbf{0}$. We have $E[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}] = (\alpha_j / \beta_j)$ and $\text{var}[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}] = \alpha_j (1+\beta_j)/\beta_j^2$ from Lemma 3;

(ii) the data is 'complete' for inference about y_j, i.e. $\mathbf{A}_j = \mathbf{e}_k$ for some *k*, and the *k*th row of $\mathbf{A}$ has the *j*th entry of one and zeros elsewhere. We have $E[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}] = x_k$ and $\text{var}[y_j \mid \mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}] = 0$ from Lemma 3.

The counterpart of equation (A1) for a normal approximation to the multivariate negative binomial distribution of **x** is

$$\Pr(\mathbf{Ay} = \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx \frac{1}{(2\pi)^{m/2} \mid \mathbf{U}_{xx} \mid^{1/2}} \exp\{-(\mathbf{x} - \mathbf{Av})^T \mathbf{U}_{xx}^{-1} (\mathbf{x} - \mathbf{Av})/2\}.$$

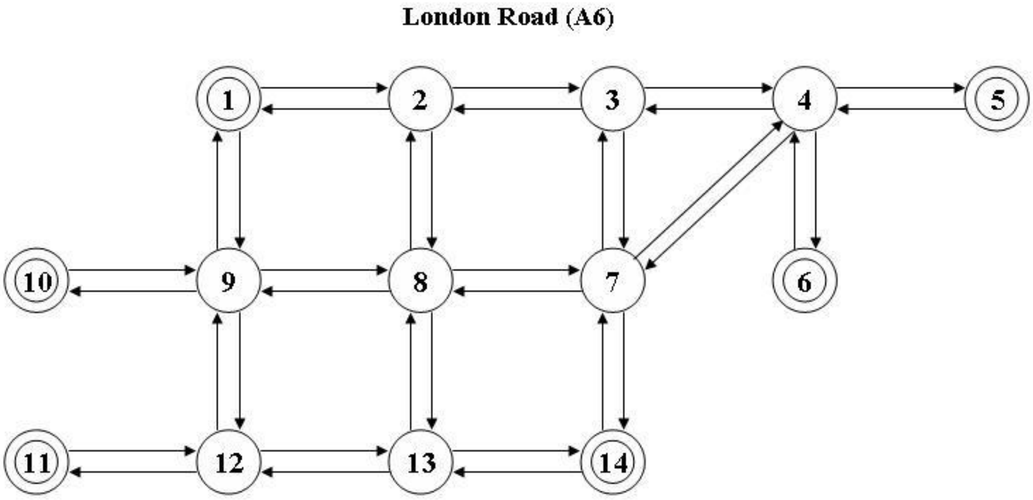

Figure 1. Abstraction of the road network in region RA of Leicester

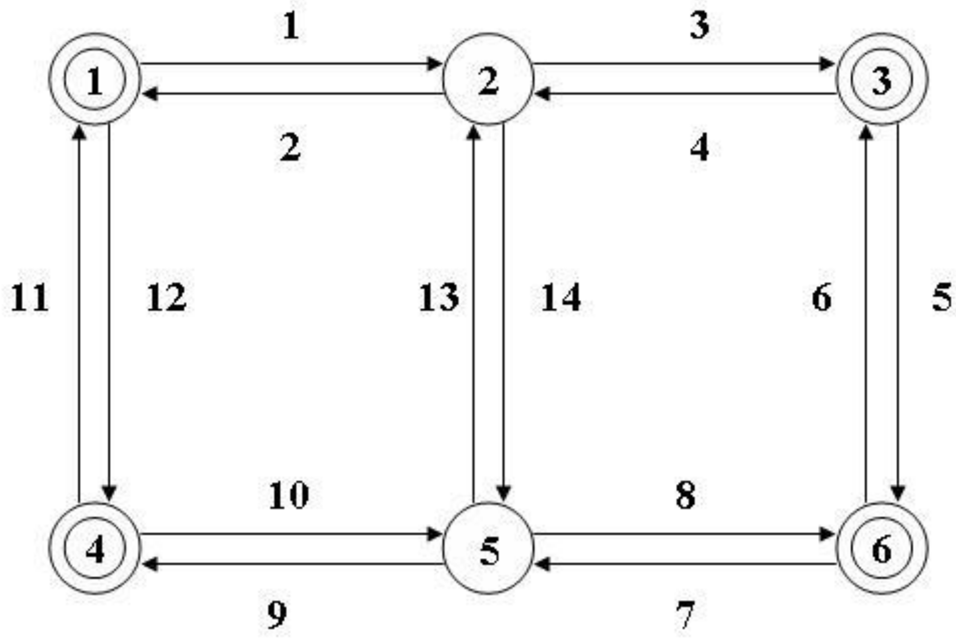

Figure 2. A simplified network topology