Book Section:

Nakayama, S, Connors, RD orcid.org/0000-0002-1696-0175 and Watling, DP orcid.org/0000-0002-6193-9121 (2009) Estimation of Parameters of Network Equilibrium Models: A Maximum Likelihood Method and Statistical Properties of Network Flow. In: Lam, WHK, Wong, SC and Lo, HK, (eds.) Transportation & Traffic Theory 2009: Golden Jubilee. Springer , pp. 39-56. ISBN 978-1-4419-0820-9

https://doi.org/10.1007/978-1-4419-0820-9_3

Estimation of Parameters of Network Equilibrium Models: A Maximum Likelihood Method and Statistical Properties of Network Flow

Shoichiro Nakayama, Kanazawa University, Japan & University of Leeds, U.K.; Richard Connors, University of Leeds, U.K.; David Watling, University of Leeds, U.K.

Abstract Estimation of the parameters in network equilibrium models, including OD matrix elements, is essential when applying the models to real-world networks. Link flow data are convenient for estimating parameters because it is relatively easy for us to obtain them. In this study, we propose a maximum likelihood method for estimating parameters of network equilibrium models using link flow data, and derive first and second derivatives of the likelihood function under the equilibrium constraint. Using the likelihood function and its derivatives, *t*-values and other statistical indices are provided to examine the confidence interval of estimated parameters and the model's goodness-of-fit. Also, we examine which conditions are needed for consistency, asymptotic efficiency, and asymptotic normality for the maximum likelihood estimators with non-I.I.D. link flow data. In order to investigate the validity and applicability, the proposed ML method is applied to a simple network and the road network in Kanazawa City, Japan.

1. Introduction

When applying traffic network equilibrium models to real-world networks, it is essential to estimate some of their parameters on both demand and supply sides. One of the most important tasks is to estimate the OD matrix, and many researchers have studied this in the past. In SUE (stochastic user equilibrium) models, we also have to calibrate a parameter associated with travel cost (or value of time) and other behavioral parameters in the random utility model if needed. In addition, parameters in the travel time functions are sometimes estimated to fit the equilibrium model using link flow data although they should normally be determined based on the relationship between link flow and travel time.

Least square error or generalized least square error has been adopted as the estimation method in most cases, especially for OD matrix estimations (Cascetta, 1984; Bell, 1991; Yang et al., 1992; Liu & Fricker, 1996; Yang et al., 2001; Nie et al., 2005). These methods enable us to estimate parameters reasonably, and are often used for network equilibrium models. The maximum likelihood method (ML method) has close relationship with the generalized squares method, indeed under some conditions, maximizing the log-likelihood function produces a particular type of generalized least squares estimator (Hazelton, 2000). Although the likelihood function is more complicated than the generalized least squares, the likelihood function can provide statistical indices for evaluating the confidence of estimated parameters and goodness-of-fit of the model, e.g. *t*-value, AIC.

Robillard (1974), Fisk (1977), and Daganzo(1977) estimated a travel cost coefficient in the logit model by means of the maximum likelihood method. Hazelton (2000) proposed a method for estimating OD matrix using maximum likelihood method. These methods are restricted to the case of uncongested networks. Anas & Kim (1990) estimated the travel cost coefficient on a congested network, but they focused on comparison and did not describe the details of the model itself.

MPEC (Mathematical Programs with Equilibrium Constraints) or bi-level optimization problems can incorporate congested networks into the maximum likelihood method. In this study, we model route choice based on random utility models and maximize the likelihood with SUE constraints. We must therefore consider the covariance of link flows arising from the flows on overlapping routes. In the past, several researchers have studied optimization problems with SUE constraints. Chen & Alfa (1991) considered discrete network design, Davis (1994) continuous network design, Liu & Fricker (1996), Yang et al. (2001) and Lo & Chan (2003) estimation of the OD matrices and the travel cost coefficient, and Ying &Yang (2005) sensitivity analysis of optimal pricing. Clark & Watling (2002) examined sensitivity analysis of probit-based SUE. These studies revealed first derivatives of the objective function. Most of the logit-based models used Dial's or modified Dial's algorithm.

In this study, we assume stochastic demands that result in stochastic link flows, and derive a joint probability distribution of link flows. Using this stochastic network equilibrium model with stochastic flows, the maximum likelihood method for a congested network is proposed. Estimation of parameters can be formulated as maximizing the likelihood with the equilibrium constraint. Second derivatives of the likelihood function are needed for calculating *t*-values, and we provide the first and second derivatives. This enables us to examine the confidence of parameters and the model itself. Also, AIC, likelihood ratio, and other indices from the likelihood are very helpful for model choice. In order to examine the validity and applicability of the proposed maximum likelihood method, we apply it to a simple network and the road network in Kanazawa City, Japan.

In travel behavior analysis, these statistical treatments with respect to the maximum likelihood method are very common. However, the maximum likelihood method is normally applied to I.I.D. data (independent, identically distributed). In case that we have the link flow data observed once, it is not I.I.D. since link flows are mutually correlated even if the data have many link flows. We examine the basic properties of the maximum likelihood estimator, consistency, asymptotic efficiency, and asymptotic normality, for the link flow data observed once. The properties are important for unbiased estimation.

2. Notation

In this section, we present some notations used in this paper.

A	=	the set of links in the network
R	=	the set of OD pairs
I	=	the set of routes in the network
I_r	=	the set of routes between the rth OD pair ($r \in R$)
q_r	=	the (realized) demand between the rth OD pair ($r \in R$)
Q_r	=	the random variable of demand between the rth OD pair
λ_r	=	the mean demand between the rth OD pair, $\sum_{r \in I_r} m_i = E[Q_r]$
x_a	=	the (realized) flow on the ath link ($a \in A$)
X_a	=	the random variable of flow on the ath link
$\mathbf{x}$	=	the vector of (realized) link flows, $(\ldots, x_a, \ldots)$
$\mathbf{X}$	=	the vector of link flow random variables, $(\ldots, X_a, \ldots)$
μ_a	=	the mean link flow on the ath link, $E[X_a]$ (= $\sum_{i \in I} \delta_{a,i}\, m_i$)
σ_a^2	=	the variance of link flow on the ath link, $Var[X_a]$
$\boldsymbol{\mu}$	=	the vector of mean link flows $(\ldots, \mu_a, \ldots)$
$\boldsymbol{\Sigma}$	=	the variance-covariance matrix of link flows
y_i	=	the flow on the ith route
Y_i	=	the random variable of flow on the ith route
$\mathbf{y}_r$	=	the vector of route flows between the rth OD pair
m_i	=	mean flow on the ith route ($i \in I$), $E[Y_i]$
$\mathbf{m}$	=	the vector of mean route flows $(\ldots, m_i, \ldots)$
p_i	=	route choice probability on the ith route
$c_a(x_a)$	=	the cost function of the ath link
$\bar{c}_a$	=	the mean cost on the ath link (= $E[c_a(X_a)]$)
$\bar{t}_i$	=	the mean cost on the ith route (= $\sum_{a \in A} \delta_{a,i}\, E[c_a(X_a)]$)
$\bar{\mathbf{t}}$	=	the vector of mean route costs $(\ldots, \bar{t}_i, \ldots)$
$\delta_{a,i}$	=	a link-route incident variable
$\boldsymbol{\Delta}$	=	a link-route incident matrix
n	=	the total number of links in the network, $\lvert A \rvert$
θ_j	=	the jth unknown parameter, which is estimated
$\boldsymbol{\theta}$	=	the vector of unknown parameters $(\ldots, \theta_j, \ldots)$

3. Stochastic Network Equilibrium Model under Poisson-Distributed Demands

In order to formulate the likelihood function, the probability density function (p.d.f.) of link flows is needed. In the past, several network equilibrium models with stochastic flows have been developed. The purpose of this study is to apply the maximum likelihood method to estimate parameters using link flow data, but not to develop a stochastic network equilibrium model with stochastic flows. In this study, we adopt the model with Poisson-distributed flows (Clark & Watling, 2005) for the sake of simplicity. Other stochastic network equilibrium models with stochastic flows can be applied to the maximum likelihood method in this study without any changes.

Assume that each OD demand follows a mutually independent Poisson distribution. Also, suppose that route choice is made independently by each driver. This means that the route flows between an OD pair follows a multinomial distribution when the demand distribution is given. Therefore, the resultant route flows between an OD pair are given by the compound of a multinomial distribution and a Poisson distribution. The route flows on the rth OD pair follow independent Poisson distributions as follows:

$$f_{\mathbf{Y}_r}(\mathbf{y}_r) = f^{mn}_{\mathbf{Y}_r}(\mathbf{y}_r | q_r)\, f^{po}_{Q_r}(q_r) \tag{1a}$$

$$= \frac{q_r!}{\prod_{i \in J_r} y_i!} \prod_{i \in I_r} \left(\frac{m_i}{\lambda_r} \right)^{y_{ij}} \frac{e^{-\lambda_r} \lambda_r^{q_r}}{q_r!} \tag{1b}$$

$$= \prod_{i=1}^{n} f^{po}_{Y_i}(y_i) \tag{1c}$$

where $f^{mn}_{\cdot}(\cdot)$ and $f^{po}_{\cdot}(\cdot)$ are the p.m.f.s (probability mass functions) of a multinomial distribution and a Poisson distribution, respectively, and $\mathbf{y}_r$ and $\mathbf{Y}_r$ are the vector of y_i and Y_i for the rth OD pair. This setting of Poisson-distributed demand and stochastic route choice is the same as Clark & Watling (2005).

As written above, $Y_i \sim Po[\mu_i]$, where $Po[\mu_i]$ denotes a Poisson distribution whose mean is μ_i. The variance of the Poisson distribution is equal to its mean, and the variance of Y_i is also m_i. When m_i is sufficiently large, $Po[m_i]$ approximates to the normal distribution, $N[m_i, m_i]$, whose mean and variance are both m_i. Also, $X_a = \sum_{i \in I} \delta_{a,i} Y_i$ and Y_i is mutually independent as Eq. 1 shows. Therefore, assume $X_a \sim N[\mu_a, \mu_a]$ in the remainder of this paper. The mean vector and the variance-covariance matrix are given by:

$$\boldsymbol{\mu} = \boldsymbol{\Delta}\mathbf{m} \tag{2}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta}\,\mathrm{diag}(\mathbf{m})\boldsymbol{\Delta}^T \tag{3}$$

where ^{T} denotes the transpose and $diag$(**m**) denotes the diagonal matrix whose diagonal component is the component of the vector **m** here and in the remainder of this paper (See Kendall & Stuart (1977), p. 375). The matrix, $\mathbf{\Delta}\mathrm{diag}(\mathbf{m})\mathbf{\Delta}^{\mathrm{T}}$, is may be singular; in this case the redundant variates should be discarded.

The p.d.f. (probability density function) of $\mathbf{X} \sim \mathrm{N}[\boldsymbol{\mu}, \mathbf{\Sigma}]$ is:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{\Sigma}|}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) \right\} \quad (4)$$

where $|\mathbf{\Sigma}|$ is a determinant of $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^{-1}$ is the inverse matrix of $\mathbf{\Sigma}$.

Assume that route choice is made based on random utility theory. In order that readers understand this paper more easily, for simplicity, assume that route choice probability is determined by the mean route travel time in this paper although it may be desirable to introduce drivers' risk-taking behavior to route choice.

$$\mathbf{p}_i(\bar{\mathbf{t}}) = \Pr\left[-\theta_1 \bar{t}_i + e_i \geq -\theta_1 \bar{t}_{i'} + e_{i'} \middle| \forall i' \in I_r, i \neq i' \right] \quad \forall i \in I_r,\ \forall r \in R \quad (5)$$

where $\bar{c}_i$ is the mean cost on the ith route, e_i is the error term, θ_1 is a positive travel cost coefficient ($\theta_1 > 0$).

Underlying the logit model is the assumption that the e_i are independent Gumbel random variables. More plausibly, it is often assumed that the route cost perceptual errors consist of link cost perceptual errors. If the link cost errors are independent normal random variables, as underlies the probit model, the route error vector $\mathbf{e} = (\ldots, e_i, \ldots)^{\mathrm{T}}$ follows a multivariate normal distribution. In this case, $\mathbf{e} \sim \mathrm{N}[\mathbf{0}, \mathbf{\Delta}\, \mathbf{\Delta}^{\mathrm{T}}]$ where $\mathbf{0}$ denotes the null vector. Alternatively, the variance-covariance matrix can be formed based on overlapping route length. $\sigma_{ii'} = \Sigma_a\, \delta_{a,i}\, \delta_{a,i'}\, d_a$ if $i \neq i'$ and otherwise, $\sigma_i^2 = \Sigma_a\, \delta_{a,i}\, d_a$ where d_a is the length of the ath link (Yai et al., 1997).

Presuppose that the parameter in Eq. 5 is an unknown parameter in this paper. Other several parameters can also be estimated. Formally, the mean demands, $\boldsymbol{\lambda}$, can be estimated in this maximum likelihood method. However, the number of OD pairs is generally much more than that of links. It is difficult to estimate OD demand stably and uniquely without other assumptions or information. For simplicity, the mean OD demands are given and fixed in this paper. The OD matrix estimation is one of the main future works.

Let $\mathbf{E}$ be an OD-route incidence matrix ($|R| \times |I|$ matrix) and $\mathbf{\Gamma} = \mathrm{diag}(\mathbf{E}^{\mathrm{T}}\boldsymbol{\lambda})$. The mean route cost is a function of mean route flows and its vector is expressed as $\bar{\mathbf{t}}(\mathbf{m})$. A Poisson-distributed network equilibrium model with random utility route choice can be formulated as the following fixed point problem:

$$\mathbf{m} = \mathbf{\Gamma}\mathbf{p}\left(\bar{\mathbf{t}}(\mathbf{m})\right) \quad (6)$$

where $\mathbf{p}(\cdot)$ is the vector function for route choice probability, $(\ldots, p_i, \ldots)^{\mathrm{T}}$. Note that $\mathbf{m}$ is a function (implicit function) of $\boldsymbol{\theta}$.

4. Maximum Likelihood Method

4.1 Single Observation

Let us now consider the case where we have a set of link flow data observed only once, that is, each observed link flow has a single value on a certain day. Let $\tilde{\mathbf{x}}$ denote the vector of observed link flows, that is, the vector of (realized) flows on observed links, and $\tilde{\mathbf{X}}$ denotes the vector of random variables of observed link flows. The distribution of observed link flows is derived as a marginal distribution of Eq. 4 and is also a multivariate normal distribution as follows:

$$\begin{aligned} f_{\tilde{X}}(\tilde{\mathbf{x}}) &= \int_{-\infty}^{\infty} f_{X}(\mathbf{x})\, d\,\tilde{\mathbf{x}}^{c} \\ &= \frac{1}{\sqrt{(2\pi)^{\tilde{n}}\left|\tilde{\boldsymbol{\Sigma}}\right|}} \exp\left\{-\frac{1}{2}(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\mu}})^{T}\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\mu}})\right\} \end{aligned} \tag{7}$$

where $\tilde{\boldsymbol{\mu}}$ is the vector of mean link flows observed, $\tilde{\boldsymbol{\Sigma}}$ is the variance-covariance matrix of observed link flows, $\tilde{n}$ is the number of observed links, and $\tilde{\mathbf{x}}^{c}$ is the vector of unobserved links. For simplicity, hereafter, every link is assumed to be observed ($\mathbf{x} = \tilde{\mathbf{x}}$, $n = \tilde{n}$), but this does not make any difference in the following discussions.

Define a log-likelihood function, $L(\boldsymbol{\theta} \mid \mathbf{x})$, as follows:

$$L(\boldsymbol{\theta}|\mathbf{x}) = \ln f_{X}(\mathbf{x}) \ . \tag{8}$$

Also, let l denote the likelihood function ($L = \ln l$), which is identical to $f_X(\mathbf{x})$ in Eq. 4.

The maximum likelihood method can be formulated as the following:

$$\max_{\boldsymbol{\theta}} \ L(\boldsymbol{\theta}|\mathbf{x}) \tag{9}$$

$$\text{s.t.} \ \ \mathbf{m} = \boldsymbol{\Gamma}\mathbf{p}(\bar{\mathbf{t}}(\mathbf{m})) \tag{10}$$

where $\boldsymbol{\theta}$ is a vector of parameters, $(\ldots,\theta_j,\ldots)^T$ ($j \in J$). In the above problem, $\boldsymbol{\theta}$ and $\mathbf{m}$ are decision variables.

4.2 Multiple Observations

When the data follows an independent identical distribution, consistency, asymptotic efficiency, and asymptotic normality of estimated parameters are guaranteed (e.g. Stuart et al., 1999).

In case of multiple independent observations, the log-likelihood function is given by:

$$L(\boldsymbol{\theta}|\mathbf{x}_1,\ldots,\mathbf{x}_V) = \sum_{v=1}^{V} \ln\ f_{X_v}(\mathbf{x}_v|\,\boldsymbol{\theta}) \tag{11}$$

where $\mathbf{x}_v$ is the link flows at the vth observation (v = 1, 2,…, V). Nevertheless, it is not easy to obtain independent link flow data observed for many days because link flows are not necessarily independent among different observations since there exist weekly, monthly, and yearly periodic cycles and there may be high correlation between observations.

If the observations are time-dependent, the log-likelihood is written as:

$$L(\boldsymbol{\theta}|\mathbf{x}_1,\ldots,\mathbf{x}_V) = \sum_{v=1}^{V} \ln\ f_{X_v}(\mathbf{x}_v|\,\boldsymbol{\theta},\mathbf{x}_1,\ldots,\mathbf{x}_{v-1}) \tag{12}$$

4.3 Derivatives of a Likelihood Function

The derivative of the likelihood function is very important for solving the problem in Eq. 9 and 10. The first derivative has been derived in the previous sensitivity analysis studies (e.g. Fiacco, 1983). In SUE, the derivatives are given using the following implicit function:

$$\mathbf{h}(\mathbf{m},\boldsymbol{\theta}) \equiv \boldsymbol{\Gamma}\mathbf{p}(\mathbf{m},\boldsymbol{\theta}) - \mathbf{m} = \mathbf{0}\ . \tag{13}$$

The first and second derivatives of $L(\Delta\mathbf{m})$ are expressed as:

$$\nabla_{\boldsymbol{\theta}} L = -(\nabla_{\mathbf{m}}\mathbf{h}^{-1}\nabla_{\boldsymbol{\theta}}\mathbf{h})^T \nabla_{\mathbf{m}} L \tag{14}$$

$$\frac{\partial^2 L}{\partial\theta_j\partial\theta_{j'}} = -\nabla_{\mathbf{m}} L^T \nabla_{\mathbf{m}}\mathbf{h}^{-1}\left(\frac{\partial^2\mathbf{h}}{\partial\theta_j\partial\theta_{j'}} + \mathbf{g}\right) + \left(\nabla_{\mathbf{m}}\mathbf{h}^{-1}\frac{\partial\mathbf{h}}{\partial\theta_j}\right)^T \nabla^2_{\mathbf{m}} L\ \nabla_{\mathbf{m}}\mathbf{h}^{-1}\frac{\partial\mathbf{h}}{\partial\theta_{j'}}\ . \tag{15}$$

where $\mathbf{g} = \left(\ldots,\frac{\partial\mathbf{m}}{\partial\theta_j}^T \nabla^2_{\mathbf{m}} h_i \frac{\partial\mathbf{m}}{\partial\theta_{j'}},\ldots\right)^T$. These derivations are derived in Appendix A.

When the parameters, $\hat{\boldsymbol{\theta}}$, are efficient, the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is $-\nabla^2_{\boldsymbol{\theta}} L^{-1}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ (Stuart et al., 1999, p. 72-75), where $\hat{\boldsymbol{\theta}}$ is the ML estimated parameter. The t-value of the jth parameter is given by:

$$\frac{\hat{\theta}_j}{\sqrt{[-\nabla^2_{\boldsymbol{\theta}} L^{-1}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}]_{jj}}} \tag{16}$$

where $[\cdot]_{jj}$ is the jj component of the matrix and $\hat{\theta}_j$ is the jth ML estimated parameter. Thus, the second derivative of L is needed for calculating t values of estimated parameters.

5. Properties of Estimated Parameters

5.1 Consistency, Asymptotic Efficiency, & Asymptotic Normality

From a statistical viewpoint, maximum likelihood estimators (ML estimators) generally have three desirable properties: consistency, asymptotic efficiency, and asymptotic normality. The mean of the estimator should be the true value of the parameter, and the smaller its variance, the better it is. This is called efficiency. The Cramer-Rao inequality provides a minimum variance (Stuart et al., 1999, p. 11). Note that ML estimators are not necessarily unbiased when the sample size is small. Unbiasedness means that E[$\breve{\theta}$] = θ_t where $\breve{\theta}$ is an estimator of θ (with θ_t its true value) and E[·] is the operator of expectation. The maximum likelihood estimator, $\hat{\theta}$, takes the true value with minimum variance in the limit as the sample size increases to infinity; this property is consistency. Furthermore, the ML estimator is asymptotically normally-distributed in the limit of large sample size. This enables us to make standard statistical tests such as t tests.

As described in section 4.2, the ML estimator has the desirable three properties if we have many independent observations. However, even though we have many observations, they may not be independent. In the case of time-dependent observations, the likelihood function is defined as Eq. 12 and the properties of ML estimators have been examined (Lehmann & Casella, 1998).

A road census is held once a year or a few years in many countries. We can use the link flow data observed once in many cases. Usually, we are able to obtain the data for many links for accurate estimation. Suppose that the number of links is sufficiently large. Each link flow is not independent and does not generally follow an independent, identical distribution. However, we guess that we can estimate parameters more accurately as the number of observed links becomes large. The above data are not time-dependent, but spatially dependent. In the time-dependent observations, the vth observation is dependent solely on the past observations and the order of observations is crucial. On the other hand, the order of spatially dependent data is not as important as that of time-dependent data, but an observation might be dependent on the other observations. We extend the understandings and findings of ML method with time-dependent observations (e.g. Lehmann & Casella, 1998) to that with spatially dependent observations. Under which conditions the ML estimators for the link flow data observed once have the properties of consistency, asymptotic efficiency and normality is examined in this study.

5.2 Local Dependency

In order to examine the properties of ML estimators with non I.I.D. link flow observation, we assume that each link flow is locally dependent. There exist many links, routes, and OD pairs in a large-scale network, and correlation between the links diminishes as distance between them become large. Assume that the distance of OD pair is finite, and each link is independent of the links that are not located within the distance of 2κ, where κ is a positive finite constant. We shall call this assumption local dependency. Note that the distance between links is defined as the distance between their centers. In case that OD demands are distributed partially or specific OD pairs has extreme demands, the local dependency should be defined based on network topology or demand distribution. For simplicity, local dependency defined by distance is adopted in this paper.

Next, we assume that each link variable has a limited variance, i.e. $\sigma_a^2 < \infty$. Therefore, $\sigma_a/\sigma_S \to 0$ as $n \to \infty$ where σ_S^2 = Var[S] and $S = \sum_a X_a$. Let us define "link density" to be the number of links whose centers are located in a unit area. Assume that the mean of link density is ρ anywhere. This means $\alpha \to \infty$ and $n/\alpha \to \rho$ as $n \to \infty$ where α is the area of the object network. Also, the area of the network can exactly be covered by a mesh of $m \times m$ squares as $n \to \infty$ where $m < \infty$. This is similar to the concepts used by Riemann integral calculus.

For the locally dependent data, consistency, asymptotic efficiency and normality of ML estimators are guaranteed. The proof is written in Appendix B. When the number of observed links is sufficiently many and the observed link flows are locally dependent, the estimated parameters are unbiased and efficient. Also, the confidence interval of the estimated parameter can be examined by the t-value due to normality of the estimated parameters.

6. Example

6.1 Simple Network

In this section, an example of a parameter estimation problem is illustrated to examine the validity of our ML method. The problem is to estimate the travel cost coefficient, θ, in the logit model using link flow data.

An example network is shown in Fig. 1. The network has two OD pairs and 4 links. The first OD pair is between Node 1 and Node 3 and the second OD pair is between Node 2 and Node 3. Both demands, Q_1 and Q_2, follow Po[2000]. The stochastic network equilibrium with Poisson distributed demands is reached, and link flows are random. The link cost functions are also written in Fig. 1. Route 1 and 2

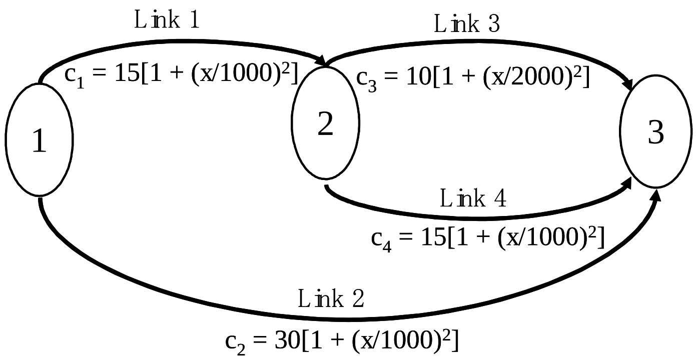

Fig. 1 Small example network

connect OD 1 and Route 3 and 4 connect OD 2. The link-route incidence matrix is given by:

$$\mathbf{\Delta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{17}$$

Note that a vehicle cannot go from Link 1 to Link 4.

The mean cost functions are derived using moment generating function (m.g.f.) as:

$$\bar{c}_1(x) = 15\left[1 + \frac{x^2 + x}{1000^2}\right] \tag{18}$$

$$\bar{c}_2(x) = 30\left[1 + \frac{x^2 + x}{1000^2}\right] \tag{19}$$

$$\bar{c}_3(x) = 10\left[1 + \frac{x^2 + x}{2000^2}\right] \tag{20}$$

$$\bar{c}_4(x) = \bar{c}_1(x) \tag{21}$$

The log-likelihood function of the example problem with single observation is given by:

$$L(\theta) = -\frac{1}{2}\left[\frac{(x_1 - m_1)^2}{m_1} + \frac{(x_2 - m_2)^2}{m_2} + \frac{(x_3 - x_1 - m_3)^2}{m_3} + \frac{(x_4 - m_4)^2}{m_4} + \ln(m_1 m_2 m_3 m_4)\right] - 2\ln(2\pi) \tag{22}$$

The true value of θ is set at 0.5. We generated 10 sets of observed link flows which are mutually independent based on Poisson distributions with the true value.

Table 1 Estimation result using all 10 observations

	m_1	m_2	m_3	m_4	θ	L	AIC
ML	1073.6 (7.00)[1]	926.4 (-7.69)[1]	1225.6 (20.54)[1]	774.4 (-25.84)[1]	0.583 (3.00)[2]	-152.93	315.86
LSE	1073.6	926.4	1227.4	772.6	0.64	–	–
True values	1073.6	926.4	1222.2	777.8	0.5	–	–
Model without θ	1073.4	926.6	1248.4	751.5	–	-158.54	325.08
Model with θ = 0	1000.0	1000.0	1000.0	1000.0	0.0	-709.81	1419.62

[1]: t value of mean flow with respect to different from that with $\theta = 0$, [2]: t value of θ

Table 1 shows the result of both ML (maximum likelihood) and LSE (least square error) methods in the case that all 10 observations are taken into account together. The likelihood function is given by Eq. 22. The table also includes the results of the model without the parameter θ and the model with $\theta = 0$. The model without θ in the table is the model with $\theta = \infty$ and its route choice is the same as Wardrop-type route choice. In the model with $\theta = 0$, the mean route flows are assigned evenly irrespective of mean route costs.

The mean route flows estimated by both ML and LSE are very close to their true values. Those estimated by ML are a little closer to their true values than those by LSE. The table shows that the estimated parameter of ML method is better than that of LSE. However, the difference between them is not statistically significant. The t-values of mean route flows and θ are all significant. Thus, the ML method proposed enables us to examine confidence intervals of parameter estimates. In the light of AIC, the model with the estimated θ has the lowest AIC of the three models (the models with the estimated θ, without θ, and with $\theta = 0$). Although the mean route flows in the model without θ are close to their true values, those with the estimated θ are much closer.

The above results show that the ML estimator is better than the LSE estimator, but the difference is not statistically significant. In the ML method, we can examine the confidence of estimated parameters and make a model choice. Also, the ML estimator is guaranteed to be unbiased if the sample size is large enough.

6.2 Kanazawa Road Network

In this section, the ML method is applied to the arterial road network in Kanazawa City, Japan. The network has 140 nodes and 472 links and is shown in Fig. 2. The number of OD pair is 1,383, and the total number of routes is 9,934. The OD matrix was given by a trip survey. The observed link flow data is a single observation

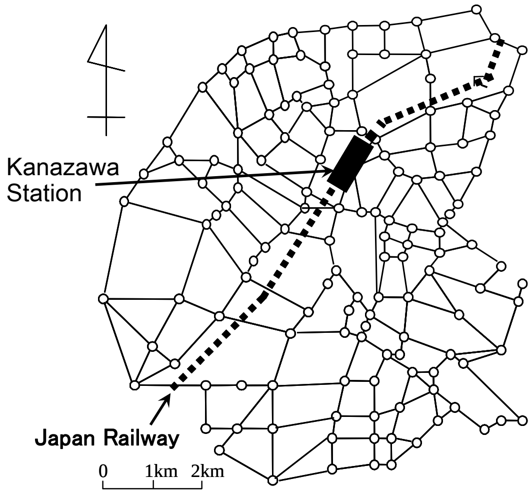

Fig. 2 Kanazawa Road Network

on a weekday and that all links are observed. The rank of the link-route incidence matrix of the network is 379. This means that flows on 93 links are calculated from the other 379 link flows. If all 472 links are treated, the inverse matrix of variance-covariance matrix for the 472 links cannot be defined because the variance-covariance matrix is singular. The 379 links are considered for estimating the parameter. Using the observed link flow data (379 link flows), the travel cost coefficient, θ_1, in the logit model is estimated. In this problem, only one parameter is estimated, and a line search algorithm with the Armijo condition is used here.

The log-likelihood function includes the p.d.f. of multivariate normal distribution, Eq. 4, and the p.d.f. has inverse matrix and determinant of variance-covariance matrix. Therefore, it is very difficult to derive the first and second derivatives of the log-likelihood function analytically. It is reasonable to use numerical differencing techniques for large-scale networks.

The estimated parameter, $\hat{\theta}$, is 0.169. The standard deviation of the estimated parameter is 0.00243, and the *t*-value is 69.7. The 99% confidence interval of θ_1 is between 0.163 and 0.175. On the other hand, the value of θ_1 estimated by the LSE method is 2.08. The LSE estimator is statistically different from the ML estimator in this Kanazawa road network example. The computation time of the LSE method is about one-fourth of that of the ML method. This is because calculating the least square error is much faster than computing the likelihood function (p.d.f. of multivariate normal distribution). The proposed ML method can be applied numerically to a real network although the computation time is slower than that of the LSE method. Also, statistical significance (or confidence interval) of the estimated parameter can be examined using the *t*-value of the estimated parameter, which derived from the second derivative of the log-likelihood function.

7. Conclusions

Estimation of parameters in network equilibrium models is essential when the models are applied to real-world networks. In this study, we proposed a method for estimating the parameters of the equilibrium models using link flow data. In the past, the least square or generalized least square method has been adopted for parameter estimation in many cases. The maximum likelihood method has close relationship with the generalized squares method. Although the likelihood function is more complicated than the generalized least squares, the likelihood function can provide statistical indices for evaluating the confidence of estimated parameters and goodness-of-fit of the model, e.g. *t*-value, AIC.

In this study, we proposed a maximum likelihood method for estimating parameters in the case of a congested network. Then, we showed that consistency, asymptotic efficiency, and asymptotic normality are guaranteed for the maximum likelihood method if many link flows are observed. In the method, t-values and other statistical indices can be provided to examine the confidence of estimated parameters and the model itself. In order to investigate the validity and applicability, the proposed ML method is applied to a simple network and the road network in Kanazawa City, Japan. It is found that the proposed ML method can be applied numerically to a real network. Also, statistical significance (or confidence interval) of the estimated parameter can also be examined using the *t*-value of the estimated parameter. As a future work, we must develop an efficient algorithm for a large-scale network.

Appendix A

Differentiate both sides of $\mathbf{h}(\mathbf{m}, \boldsymbol{\theta}) = \mathbf{0}$ with respect to θ_j, then

$$\frac{\partial \mathbf{h}}{\partial \theta_j} + \nabla_{\mathbf{m}} \mathbf{h} \frac{\partial \mathbf{m}}{\partial \theta_j} = \mathbf{0}.$$

Therefore,

$$\frac{\partial \mathbf{m}}{\partial \theta_j} = -\nabla_{\mathbf{m}} \mathbf{h}^{-1} \frac{\partial \mathbf{h}}{\partial \theta_j}.$$

The above equation is expressed in the vector form by:

$$\nabla_{\boldsymbol{\theta}} \mathbf{m} = -\nabla_{\mathbf{m}} \mathbf{h}^{-1} \nabla_{\boldsymbol{\theta}} \mathbf{h}.$$

We obtain the following first derivative:

$$\nabla_{\boldsymbol{\theta}} \mathrm{L} = \nabla_{\boldsymbol{\theta}} \mathbf{m}^{\mathrm{T}} \nabla_{\mathbf{m}} \mathrm{L} = -(\nabla_{\mathbf{m}} \mathbf{h}^{-1} \nabla_{\boldsymbol{\theta}} \mathbf{h})^{\mathrm{T}} \nabla_{\mathbf{m}} \mathrm{L}$$

Let us consider the second derivative. Differentiate $\frac{\partial \mathbf{h}}{\partial \theta_j} + \nabla_{\mathbf{m}} \mathbf{h} \frac{\partial \mathbf{m}}{\partial \theta_j} = \mathbf{0}$ with respect to $\theta_{j'}$,

$$\frac{\partial \mathbf{h}}{\partial \theta_j \partial \theta_{j'}} + \frac{\partial \nabla_{\mathbf{m}} \mathbf{h}}{\partial \theta_{j'}} \frac{\partial \mathbf{m}}{\partial \theta_j} + \nabla_{\mathbf{m}} \mathbf{h} \frac{\partial \mathbf{m}}{\partial \theta_j \partial \theta_{j'}} = \mathbf{0} .$$

Let $\mathbf{g}$ denote $\frac{\partial \nabla_{\mathbf{m}} \mathbf{h}}{\partial \theta_{j'}} \frac{\partial \mathbf{m}}{\partial \theta_j}$. $\mathbf{g} = \left(\ldots, \frac{\partial \mathbf{m}}{\partial \theta_j}^{T} \nabla^2_{\mathbf{m}} h_i \frac{\partial \mathbf{m}}{\partial \theta_{j'}}, \ldots \right)^{T}$ because $\frac{\partial \nabla_{\mathbf{m}} h_i}{\partial \theta_{j'}} = \nabla^2_{\mathbf{m}} h_i \frac{\partial \mathbf{m}}{\partial \theta_{j'}}$.

Accordingly,

$$\frac{\partial \mathbf{m}}{\partial \theta_j \partial \theta_{j'}} = -\nabla_{\mathbf{m}} \mathbf{h}^{-1} \left(\frac{\partial \mathbf{h}}{\partial \theta_j \partial \theta_{j'}} + \mathbf{g} \right)$$

From the derivative of composite function,

$$\frac{\partial^2 L}{\partial \theta_j \partial \theta_{j'}} = \frac{\partial}{\partial \theta_{j'}} \left(\nabla_{\mathbf{m}} L^{T} \frac{\partial \mathbf{m}}{\partial \theta_j} \right) = \nabla_{\mathbf{m}} L^{T} \frac{\partial^2 \mathbf{m}}{\partial \theta_j \partial \theta_{j'}} + \left(\nabla^2_{\mathbf{m}} L \frac{\partial \mathbf{m}}{\partial \theta_{j'}} \right)^{T} \frac{\partial \mathbf{m}}{\partial \theta_j} .$$

Therefore, we obtain the second derivative as follows:

$$\frac{\partial^2 L}{\partial \theta_j \partial \theta_{j'}} = -\nabla_{\mathbf{m}} L^{T} \nabla_{\mathbf{m}} \mathbf{h}^{-1} \left(\frac{\partial^2 \mathbf{h}}{\partial \theta_j \partial \theta_{j'}} + \mathbf{g} \right) + \left(\nabla_{\mathbf{m}} \mathbf{h}^{-1} \frac{\partial \mathbf{h}}{\partial \theta_j} \right)^{T} \nabla^2_{\mathbf{m}} L \, \nabla_{\mathbf{m}} \mathbf{h}^{-1} \frac{\partial \mathbf{h}}{\partial \theta_{j'}} .$$

Appendix B

B.1 Weak Law of Large Numbers for Locally Dependent Variates

In order to examine the ML estimator for spatially dependent data, we need the weak law of large numbers (WLLN) and the central limit theorem (CLT) for spatially dependent random variables. In this section, we prove the weak law of large numbers under the assumption of local dependency.

From Chebyshev's inequality,

$$\Pr\left[\left| \overline{X} - \mu \right| > \varepsilon \right] \le \frac{1}{\varepsilon^2} \mathrm{Var}\left[\overline{X} \right] = \frac{\mathrm{Var}[S]}{n^2 \varepsilon^2}$$

where $\overline{X} = \sum_{a \in A} X_a / n$ (= S/n), $\mu \equiv \sum_{a \in A} \mu_a / n$, $S = \sum_{a \in A} X_a$, $\mathrm{Var}[S] = \sum_{a \in A} \sum_{a' \in A} \sigma_{aa'}$, and $\sigma_{aa'}$ is:

$$\sigma_{aa'} = \begin{cases} \mathrm{Var}[X_a] & \text{if } a = a' \\ \mathrm{Cov}[X_a, X_{a'}] & \text{otherwise} \end{cases}$$

Because of the local dependency as described above, there exists $\gamma_a < \infty$ satisfying $|\sum_{a\in A} \sigma_{aa'}| < \gamma_a$. Let γ be $\max[|\gamma_a|;\forall a \in A]$ ($\gamma < \infty$) where max is an operator of taking a maximal value. Therefore, $\mathrm{Var}[S] < \sum_{a\in A} \gamma_a \le n\,\gamma$, and

$$\Pr\left[\left|\bar{X} - \mu\right| > \varepsilon\right] < \frac{\gamma}{n\varepsilon^2}$$

As $n \to \infty$, $\Pr[\bar{X}_n - \mu > \varepsilon] = 0$ ($\forall \varepsilon > 0$). That is, $\bar{X}$ converged in probability to μ. We shall call this the weak law of large numbers for locally dependent random variables.

B.2 Central Limit Theorem for Locally Dependent Variates

The central limit theorem (CLT) usually requires independence of random variables. Here we examine the CLT for locally dependent variables.

Choose and fix β ($0 < \beta < 1/2$) and let $\eta = [n^{\beta}]$ and $\nu = [n/\eta]$, where $[x]$ denotes the largest integer $\le x$.

We have $\eta \le n^{\beta}$. It is easily seen that $\eta = O(n^{\beta})$, and $\nu = O(n^{1-\beta})$.

Let m^2 be α/ν where α is the area of the network and m has been quoted in section 5.2. As $\alpha \to \infty$ (i.e. $n \to \infty$), α can nearly be divided into ν pieces of $m\times m$ square as described in section 5.2. Namely, $\cup_k B_k \to \alpha$ as $n \to \infty$ where B_k is the kth $m \times m$ square which comprises the area of the network and $B_k \cap B_{k'} = \phi$ ($k \ne k'$). Note that B_k means the set of links within it as well as the geometric square.

Consider a $(m - 2\kappa)\times(m - 2\kappa)$ square. Let C_k denote the kth $(m - 2\kappa)\times(m - 2\kappa)$ square whose center is the same center of B_k.

Now, let U_k and $\bar{U}_k$ denote the sum of link variables within C_k and the sum of links within B_k but out of C_k, respectively. Fig. 3 illustrates them. In the figure, the grey square is C_k and the sum of links in C_k is U_k. Also, $U = \sum_k U_k$ and $\bar{U} = \sum_k \bar{U}_k$. Then, $S = U + \bar{U}$.

The proof consists of 1) U has a limiting normal distribution and 2) $\bar{U} \to 0$ in probability.

At first, let's prove 2), and consider the variance of $\bar{U}$.

$$\sigma_{\bar{U}}^2 = \sum_{a\in A_{\bar{U}}} \left(\sigma_a^2 + \sum_{a'\in A_{a'}} \sigma_{aa'} \right)$$

where $\sigma_{\bar{U}}^2$ is the variance of $\bar{U}$, and $A_{\bar{U}}$ is the set of links which consist of $\bar{U}$, and $A_{a'}$ is the set of links on which the a′th link is dependent.

$$\mathrm{Var}[\overline{\mathrm{U}}/\sigma_{\mathrm{S}}] = \mathrm{Var}[\overline{\mathrm{U}}]/\sigma_{\mathrm{S}}{}^2 = \frac{\sigma_{\overline{\mathrm{U}}}^2}{\sigma_{\mathrm{S}}^2}.$$

Let λ be $\max[\sigma^2_{\overline{\mathrm{U}}_\mathrm{k}}, \sigma_{\overline{\mathrm{U}}_\mathrm{k}\overline{\mathrm{U}}_{\mathrm{k}'}} \,|\forall \mathrm{k}, \forall \mathrm{k}']$. The number of links in $\mathrm{A}_{\overline{\mathrm{U}}}$ is $\nu\,[\mathrm{m}^2 - (\mathrm{m} - 2\kappa)^2]\,\rho = 4\,\kappa\,\nu\rho\,(\mathrm{m} - \kappa)$ ($\equiv \mathrm{n}_{\overline{\mathrm{U}}}$) as $\mathrm{n} \to \infty$ where ν is the number of $\overline{\mathrm{U}}_\mathrm{k}$ because of the link density $\rho = \mathrm{n}/\alpha$. As written above, $\sigma_{\overline{\mathrm{U}}}^2 \le \mathrm{n}_{\overline{\mathrm{U}}}\lambda(1+\tau) = 4\kappa\nu\lambda\rho(\mathrm{m}-\kappa)(1+\tau)$ where τ is $\max[\mathrm{n}_{\overline{\mathrm{U}}_\mathrm{k}} \,|\forall \mathrm{k}]$ and $\mathrm{n}_{\overline{\mathrm{U}}_\mathrm{k}}$ is the number of links which is dependent of $\overline{\mathrm{U}}_\mathrm{k}$. So, the order of $\sigma_{\overline{\mathrm{U}}}^2$ is:

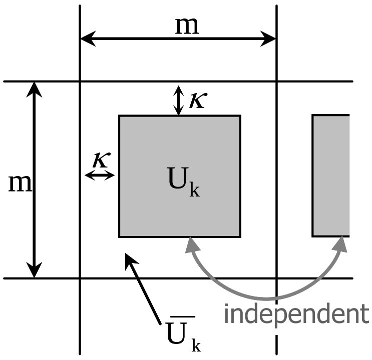

Fig. 3 Explanation of *U*

$$\sigma_{\overline{\mathrm{U}}}^2 = \mathrm{O}(\mathrm{n}^{\beta/2})\,\mathrm{O}(\mathrm{n}^{1-\beta}) = \mathrm{O}(\mathrm{n}^{1-\beta/2})$$

where because κ, λ, and τ are finite fixed values and $\mathrm{m} = \mathrm{O}(\mathrm{n}^{\beta/2})$ and $\nu = \mathrm{O}(\mathrm{n}^{1-\beta})$.

Thus, the order of $\sigma_{\overline{\mathrm{U}}}^2$ is $\mathrm{n}^{1-\beta/2}$ while the order of σ_{S}^2 is n. Therefore, $\sigma_{\overline{\mathrm{U}}}^2 / \sigma_{\mathrm{S}}^2 \to 0$ as $\mathrm{n} \to \infty$ because $\beta < 1/2$.

From Chebyshev's inequality, $\Pr\left[\left|\overline{\mathrm{U}} - \mu_{\overline{\mathrm{U}}}\right| > \sigma_{\mathrm{S}}\right] \le \sigma_{\overline{\mathrm{U}}}^2 / \sigma_{\mathrm{S}}^2$. As written above, $\sigma_{\overline{\mathrm{U}}}^2 / \sigma_{\mathrm{S}}^2 \to 0$ as $\mathrm{n} \to \infty$, and $\overline{\mathrm{U}} \to \mu_{\overline{\mathrm{U}}}$ in probability (as $\mathrm{n} \to \infty$) where $\mu_{\overline{\mathrm{U}}}$ is the mean of $\overline{\mathrm{U}}$.

Next, let's move to 1). The number of links which consist of U is $\mathrm{n_U} = \nu\,(\mathrm{m} - 2\kappa)^2\rho$ as $\mathrm{n} \to \infty$. The order of $\mathrm{n_U}$ is n, and $\mathrm{n_U} \to \mathrm{n}$ as $\mathrm{n} \to \infty$ because of $\mathrm{m} = \mathrm{O}(\mathrm{n}^{\beta/2})$ and $\nu = \mathrm{O}(\mathrm{n}^{1-\beta})$. This means that $\mathrm{U} \to \mathrm{S}$ as $\mathrm{n} \to \infty$. Namely, $\sigma_{\mathrm{U}}^2 \to \sigma_{\mathrm{S}}^2$ as $\mathrm{n} \to \infty$. Therefore,

$$\frac{\mathrm{S} - \mu_{\mathrm{S}}}{\sigma_{\mathrm{S}}} = \frac{\mathrm{U} - \mu_{\mathrm{U}}}{\sigma_{\mathrm{S}}} + \frac{\overline{\mathrm{U}} - \mu_{\overline{\mathrm{U}}}}{\sigma_{\mathrm{S}}} \to \frac{\mathrm{U} - \mu_{\mathrm{U}}}{\sigma_{\mathrm{U}}}$$

as $\mathrm{n} \to \infty$ because $(\overline{\mathrm{U}} - \mu_{\overline{\mathrm{U}}})/\sigma_{\mathrm{S}} \to 0$ from 2). $(\mathrm{U}-\mu_{\mathrm{U}})/\sigma_{\mathrm{U}}$ has a limiting standard normal distribution because $\mathrm{U_k}$ is mutually independent and (Lindeberg's) CLT (e.g. Kendal & Stuart, 1977, p. 206-207) can be applied (since $\sigma_{\mathrm{j}}/\sigma_{\mathrm{S}} \to 0$ as $\mathrm{n} \to \infty$). Consequently, $(\mathrm{S} - \mu_{\mathrm{S}})/\sigma_{\mathrm{S}}$ follows a standard normal distribution, and $\mathrm{S} \sim \mathrm{N}[\mu_{\mathrm{S}}, \sigma_{\mathrm{S}}{}^2]$. Thus, the CLT for local dependent variables has been proven.

We will consider $\nabla_{\boldsymbol{\theta}} \ln f_{\mathrm{X_a}}(\mathrm{x_a}|\boldsymbol{\theta})$ as a vector of random variables in the next section. Let us extend the above to the multivariate CLT. The vector of random variables on the *a*th link is denoted by $\mathbf{y}_\mathrm{a}$. Let $\mathbf{S}_\mathbf{y}$ be $\Sigma_\mathrm{a}\, \mathbf{y}_\mathrm{a}$ and $\mathbf{I}$ be the variance-covariance matrix of $\mathbf{y}$. In the next section, $\mathbf{S}_\mathbf{y} = \nabla_{\boldsymbol{\theta}}\, \mathrm{L}$.

Let $\mathrm{Z_a} = \mathbf{d}^\mathrm{T}\, \mathbf{Y}_\mathrm{a}$ and $\mathrm{S_z} = \Sigma_\mathrm{a}\, \mathrm{Z_a}$. The p.d.f. of $\mathbf{y} \propto \mathbf{y}^\mathrm{T}\, \mathbf{I}^{-1}\, \mathbf{y} = (\mathbf{d}^\mathrm{T}\, \mathbf{z})^\mathrm{T}\, \mathbf{I}^{-1}\, \mathbf{d}^\mathrm{T}\, \mathbf{z} = \mathbf{z}^\mathrm{T}\, (\mathbf{d}^\mathrm{T}\, \mathbf{I}\, \mathbf{d})^{-1}\, \mathbf{z}$, so $\mathrm{Var}[\mathrm{S_z}] = \mathbf{d}^\mathrm{T}\, \mathbf{I}\, \mathbf{d}$. By the above CLT, $\mathrm{S_z} - \mathrm{E}[\mathrm{S_z}] \sim \mathrm{N}[0, \mathbf{d}^\mathrm{T}\, \mathbf{I}\, \mathbf{d}]$ as $\mathrm{n} \to \infty$. By the Cramér-Wold devise (Durrett, 1991), $\mathbf{S}_\mathbf{y}$ is multinormally distributed and its the variance-covariance matrix is $\mathbf{I}$.

B.3 Properties of ML Estimators for Local Dependent Data

Using the above WLLN and CLT, we show consistency, asymptotic efficiency and normality of each ML estimator of parameters for locally dependent data. The outline is almost the same as those for I.I.D. data except usage of WLLN and CLT for locally dependent data, and we illustrate them briefly.

The likelihood function, l, is the joint density function of the observations, and evidently,

$$\int\cdots\int \mathrm{l}\,\mathrm{d}x_1\cdots\mathrm{d}x_n = 1\,. \tag{b1}$$

Now suppose that the first two derivatives of l with respect to θ_j exist for all θ_j, and the operations of differentiation and integration on its left-hand side. We obtain:

$$\mathrm{E}\left[\frac{\partial \log \mathrm{l}}{\partial\theta_j}\right] = \mathrm{E}\left[\frac{\partial \mathrm{L}}{\partial\theta_j}\right] = \int\cdots\int\left(\frac{1}{\mathrm{l}}\frac{\partial \mathrm{l}}{\partial\theta_j}\right)\mathrm{l}\,\mathrm{d}x_1\cdots\mathrm{d}x_n = 0\,. \tag{b2}$$

If we differentiate the above equation, we have:

$$\int\cdots\int\left\{\frac{\partial \mathrm{L}}{\partial\theta_j}\frac{\partial \mathrm{L}}{\partial\theta_{j'}} + \frac{\partial^2 \mathrm{L}}{\partial\theta_j\partial\theta_{j'}}\right\}\mathrm{l}\,\mathrm{d}x_1\cdots\mathrm{d}x_n = \mathrm{E}\left[\frac{\partial \mathrm{L}}{\partial\theta_j}\frac{\partial \mathrm{L}}{\partial\theta_{j'}}\right] + \mathrm{E}\left[\frac{\partial^2 \mathrm{L}}{\partial\theta_j\partial\theta_{j'}}\right] = 0\,.$$

Using Taylor's theorem, we have

$$\nabla_{\boldsymbol\theta}\mathrm{L}\big|_{\boldsymbol\theta=\hat{\boldsymbol\theta}} = \nabla_{\boldsymbol\theta}\mathrm{L}\big|_{\boldsymbol\theta=\hat{\boldsymbol\theta}} + \nabla^2_{\boldsymbol\theta}\mathrm{L}\big|_{\boldsymbol\theta=\boldsymbol\theta^*}(\hat{\boldsymbol\theta}-\boldsymbol\theta_t) = \mathbf{0}\,.$$

where $\boldsymbol\theta^*$ is some value between $\hat{\boldsymbol\theta}$ and $\boldsymbol\theta_t$.

$\hat{\boldsymbol\theta}$ is a ML estimator, and $\nabla_{\boldsymbol\theta}\mathrm{L}\big|_{\theta=\hat\theta} = \mathbf{0}$. Also, $\partial\mathrm{L}/\partial\theta_j\big|_{\theta_j=\theta_{tj}} \to 0$ in probability from Eq. (22) where θ_{tj} is the jth component of $\boldsymbol\theta_t$ (the vector of true values). Since $\partial^2\mathrm{L}/\partial\theta_j{}^2 \to -\mathrm{E}[(\partial\mathrm{L}/\partial\theta_j)^2] < 0$ as $n\to\infty$ by the WLLN for locally dependent variables, $(\hat{\boldsymbol\theta}-\boldsymbol\theta_t)\to\mathbf{0}$ from Eq. b1. This means consistency of the ML estimator.

Let $\mathbf{I}$ denote the Fisher's information matrix.

$$[\mathbf{I}]_{jj'} = -\mathrm{E}\left[\frac{\partial^2\mathrm{L}}{\partial\theta_j\partial\theta_{j'}}\right] = \mathrm{E}\left[\frac{\partial\mathrm{L}}{\partial\theta_j}\frac{\partial\mathrm{L}}{\partial\theta_{j'}}\right]$$

where $\mathbf{y} = \nabla_{\boldsymbol\theta}\mathrm{L}\big|_{\theta=\hat\theta}$, and. and denote

$\dfrac{\partial^2\mathrm{L}}{\partial\theta_j\partial\theta_{j'}} = \sum_{i=1}^{n}\dfrac{\partial^2\ln f_i}{\partial\theta_j\partial\theta_{j'}} \to \mathrm{E}\left[\dfrac{\partial^2\mathrm{L}}{\partial\theta_j\partial\theta_{j'}}\right]$ (as $n\to\infty$) due to the WLLN for locally dependent variables. Therefore, from $\nabla_{\boldsymbol\theta}\mathrm{L}\big|_{\theta=\hat\theta} = \mathbf{0}$, we rewrite Eq. b1 in the form:

$$\hat{\boldsymbol\theta}-\boldsymbol\theta_t = \mathbf{I}^{-1}\mathbf{y}$$

where $\mathbf{y} = \nabla_{\boldsymbol\theta}\mathrm{L}\big|_{\theta=\hat\theta}$.

Eq. b2 means that $\mathbf{I}$ is the variance-covariance matrix of $\mathbf{y}$. By the CLT for locally dependent variables, $\mathbf{y} \sim N[\mathbf{0}, \mathbf{I}]$. The p.d.f. of $\mathbf{y}$ is proportional to $\exp(-\mathbf{y}^T \mathbf{I}^{-1} \mathbf{y}/2)$, and $\exp(-\mathbf{y}^T \mathbf{I}^{-1} \mathbf{y}/2) = \exp[-(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_t)^T \mathbf{I} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_t)/2]$. Accordingly,

$$\hat{\boldsymbol{\theta}} \sim N[\boldsymbol{\theta}_t, \mathbf{I}^{-1}].$$

This mean that the ML estimator has the minimum variance and is a normal variate as $n \to \infty$. Namely, asymptotic efficiency and normality have been proven.

Acknowledgments I would like to express my gratitude to Prof. Jun-ichi Takayama (Kanazawa University) for providing valuable comments and advices. I am also grateful to Mr. Tomonari Anaguchi (Kanazawa University) for his computational support.

References

Anas, A. & Kim, I. (1990) Network Loading versus Equilibrium Estimation of the Stochastic Route Choice Model: Maximum Likelihood and Least Squares Revisited. *Journal of Regional Science*, 30, 89-103.

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

Cascetta, E. (1984) Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator. *Transportation Research*, 18B, 289-299.

Chen, M. & Alfa, A. (1991) A Network Design Algorithm Using a Stochastic Incremental Traffic Assignment Approach. *Transportation Science*, 25, 215-224.

Clark, S.D & Watling, D.P. (2002) Sensitivity Analysis of the Probit-Based Stochastic User Equilibrium Assignment Model. *Transportation Research*, 36B, 617-635.

Clark, S.D & Watling, D.P. (2005) Modelling Network Travel Time Reliability under Stochastic Demand. *Transportation Research*, 39B, 119-140.

Daganzo, C.F. (1977) Some Statistical Problems in connection with Traffic Assignment. *Transportation Research*, 11, 385-389.

Davis, G.A. (1994) Exact Local Solution of the Continuous Network Design Problem via Stochastic User Equilibrium Assignment. *Transportation Research*, 28B, 61-75.

Durrett, R. (1991) *Probability: Theory and Examples*. Pacific Grove, CA.: Wadsworth and Brooks/Cole.

Fiacco, A. (1983) *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. New York: Academic Press.

Fisk, C. (1977) Note on the Maximum Likelihood calibration on Dial's Assignment Method. *Transportation Research*, 11, 67-68.

Kendall, M. & Stuart, A. (1977) *The Advanced Theory of Statistics, Vol. 1, Distribution Theory*. 4th ed., London: Charles Griffin,.

Hazelton, M.L. (2000) Estimation of Origin-Destination Matrices from Link Flows under Uncongested Networks. *Transportation Research*, 34B, 549-566.

Lehmann, E.L. & Casella, G. (1998) *Theory of Point Estimation*. Berlin: Springer.

Liu, S.S. & Fricker, J.D. (1996) Estimation of a Trip Table and the θ Parameter in a Stochastic Network. *Transportation Research*, 30A, 287-305.

Lo, H.P. & Chan, C.P. (2003) Simultaneous estimation of an origin-destination matrix and link choice proportions using traffic counts. *Transportation Research*, 37B, 771-788.

Nie, Y., Zhang, H.M., Recker W.W. (2005) Inferring Origin-Destination Trip Matrices with a Decoupled GLS Path Flow Estimator. *Transportation Research*, 39B, 497-518.

Robillard, P. (1974) Calibration of Dial's Assignment Method. *Transportation Science*, 8, 117-125.

Stuart, A, Ord, J.K., Arnold, S. (1999) *Kendall's Advanced Theory of Statistics, Vol. 2A, 6th Ed.* London: Arnold Publisher.

Yai, T., Iwakura, S. Morichi, S. (1997) Multinomial Probit with Structured Covariance for Route Choice Behavior. *Transportation Research*, 31B, 195-207.

Yang, H., Sasaki, T. and Iida, Y. (1992) Estimation of origin-destination matrices from link traffic counts on congested networks, *Transportation Research*, Vol. 26B, pp. 417-434.

Yang, H., Meng, Q., Bell, M.G.H (2001) Simultaneous Estimation of the Origin-Destination Matrices and Travel-Cost Coefficient for Congested Networks in a Stochastic User Equilibrium. *Transportation Science*, 35, 107-123.

Ying, J.Q. & Yang, H. (2005) Sensitivity Analysis of Stochastic User Equilibrium Flows in a Bi-Modal Network with Application to Optimal Pricing. *Transportation Research*, 39B, 769-795.