Statistical Inference for Origin-Destination Matrices

Martin Hazelton
m.hazelton@massey.ac.nz
Massey University, New Zealand

Presented to the Hong Kong Society for Transportation Studies, Feb. 2012
Outline

1 Fundamentals
   - Introduction to the Estimation Problem
   - Model Building
   - What are We Trying to Estimate?
   - Statistical Models

2 Indeterminacy

3 Methods of Estimation
   - Maximum Likelihood Estimation
   - Method of Moments
   - Bayesian Methods

4 Accounting for Measurement Error

5 Incorporating Sporadic Routing Information

6 End Matter
   - Acknowledgements
   - References
A Loose Statement of the Problem

Given traffic counts on links network, estimate the rate of travel between each pair of (origin/destination) nodes.
Static or Dynamic?

- Assume that (daily) traffic counts are totals over some observational window.
- Aim will be to estimate origin-destination (OD) travel rates over same window.
- This is typically viewed as the static OD estimation problem.
- Critically difference from within-day dynamic OD estimation problem:
  - **Within-day Dynamic**: Journeys continue through several observational windows.
  - **Static**: Journeys assumed complete in a single observational window.
Static or Dynamic?

- Assume that (daily) traffic counts are totals over some observational window.
- Aim will be to estimate origin-destination (OD) travel rates over same window.
- This is typically viewed as the static OD estimation problem.
- Critically difference from within-day dynamic OD estimation problem:
  - Within-day Dynamic: Journeys continue through several observational windows.
  - Static: Journeys assumed complete in a single observational window.

The focus here is ‘static’ OD matrix estimation.
Sequences of Link Counts

- Recall: traffic counts over fixed window (e.g. 4.00-5.00pm).
- We can usually collect such counts over a sequence of days.

**London Road (link 5)**

<table>
<thead>
<tr>
<th>Day</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>950</td>
</tr>
<tr>
<td>10</td>
<td>1150</td>
</tr>
<tr>
<td>15</td>
<td>1150</td>
</tr>
</tbody>
</table>

**Waterloo Way (link 27)**

<table>
<thead>
<tr>
<th>Day</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>550</td>
</tr>
<tr>
<td>10</td>
<td>650</td>
</tr>
<tr>
<td>15</td>
<td>750</td>
</tr>
</tbody>
</table>
Variables

\[
\begin{align*}
\mathbf{y} &= (y_1, \ldots, y_N)^T \quad \text{link (arc) flows} \\
\mathbf{x} &= (x_1, \ldots, x_M)^T \quad \text{route (path) flows} \\
\mathbf{u} &= (u_1, \ldots, u_L)^T \quad \text{OD flows}
\end{align*}
\]

- Only link counts \( \mathbf{x} \) are directly observed.
- Superscript added when necessary to indicate time.
- Hence \( \{ \mathbf{y}^t : t = 1, \ldots, n \} = \{ \mathbf{y}^1, \ldots, \mathbf{y}^n \} \) denotes sequence of link counts over \( n \) days.
- Note \( y^t_i \) and \( u^t_j \) are \textbf{realized numbers} of vehicles on route \( i \) and OD pair \( j \) on day \( t \).
Relationships Between Variables

Relationship between link counts and route flows

\[ y = Ax \]

- \( A = (a_{ij}) \) is link-route incidence matrix.
- \( a_{ij} = 1 \) if link \( i \) on route \( j \), \( 0 \) otherwise.

Relationship between route flows and OD flows

\[ u = Bx \]

- \( B = (b_{ij}) \) is OD pair-route incidence matrix.
- \( b_{ij} = 1 \) if route \( j \) services OD pair \( i \), \( 0 \) otherwise.
Relationships Between Variables

Relationship between link counts and route flows

\[ y = Ax \]

- \( A = (a_{ij}) \) is link-route incidence matrix.
  - \( a_{ij} = 1 \) if link \( i \) on route \( j \), \( 0 \) otherwise.

Relationship between route flows and OD flows

\[ u = Bx \]

- \( B = (b_{ij}) \) is OD pair-route incidence matrix.
  - \( b_{ij} = 1 \) if route \( j \) services OD pair \( i \), \( 0 \) otherwise.

System is inherently linear (just a counting exercise).
Random Variables

- Link counts vary in a (partially) haphazard manner from day to day.
- Therefore natural to model $y^1, \ldots, y^n$ as random variables.
Random Variables

- Link counts vary in a (partially) haphazard manner from day to day.
- Therefore natural to model $y^1, \ldots, y^n$ as random variables.
- Follows that $x^1, \ldots, x^n$ and $u^1, \ldots, u^n$ are random variables.
- Recognizes that actual OD flow varies from day to day.
Model Parameters: Mean Values

**Mean route flow vector**

Define $\lambda = \mathbb{E}[x]$ to be mean route flow.
Model Parameters: Mean Values

Mean route flow vector
Define $\lambda = \mathbb{E}[x]$ to be mean route flow.

Hence ...

Mean link count vector
$\mathbb{E}[y] = A\lambda$

Mean OD flow vector
$\mu = \mathbb{E}[u] = B\lambda$
What are We Trying to Estimate?

OD flows can be described by:
What are We Trying to Estimate?

OD flows can be described by:

- $u^1, \ldots, u^n$: sequence realized OD flows.
What are We Trying to Estimate?

OD flows can be described by:

- $u^1, \ldots, u^n$: sequence realized OD flows.
- $\bar{u} = n^{-1} \sum_{t=1}^{n} u^t$: sample mean OD flow vector.
What are We Trying to Estimate?

OD flows can be described by:

- \( u^1, \ldots, u^n \): sequence realized OD flows.
- \( \bar{u} = n^{-1} \sum_{t=1}^{n} u^t \): sample mean OD flow vector.
- \( \mu = \mathbb{E}[u] \): ‘population’ mean OD flow vector.
What are We Trying to Estimate?

OD flows can be described by:

- \(u^1, \ldots, u^n\): sequence realized OD flows.
- \(\bar{u} = n^{-1} \sum_{t=1}^{n} u^t\): sample mean OD flow vector.
- \(\mu = \mathbb{E}[u]\): ‘population’ mean OD flow vector.
- \(u^*\): actual OD flow that will arise tomorrow.
What are We Trying to Estimate?

OD flows can be described by:

- \( u^1, \ldots, u^n \): sequence realized OD flows.
- \( \bar{u} = n^{-1} \sum_{t=1}^{n} u^t \): sample mean OD flow vector.
- \( \mu = \mathbb{E}[u] \): ‘population’ mean OD flow vector.
- \( u^* \): actual OD flow that will arise tomorrow.

Which of these do we really want to estimate?
Taxonomy of OD Estimation Problems

**Reconstruction**
Aim is to estimate $u^t$ for some past day $t$.

**Estimation**
Aim is to estimate $\mu$ (classical statistical approach)

**Prediction**
Aim is to estimate actual future OD flow $u^*$. 

Taxonomy of OD Estimation Problems

Reconstruction
Aim is to estimate $u^t$ for some past day $t$.

Estimation
Aim is to estimate $\mu$ (classical statistical approach)

Prediction
Aim is to estimate actual future OD flow $u^*$.

- Optimal predictor of $u^*$ is $\mu$ (under regular conditions).
- Follows that estimation problem usually principal focus.

What are We Trying to Estimate?

Time Varying Mean OD Flows

- Have assumed implicitly that $E[u_t] = \mu$ does not vary with $t$.
- What happens if it does? I.e. $E[u_t] = \mu^t$.

Possible approaches...

1. **Decompose into Daily Problems**
   - Seems to lose information if $\{\mu^t\}$ do not vary too much.
   - Standard statistical theory not applicable.

2. **Model Variation in $\mu^t$**
   - Represent $\mu^t = \mu(\psi, t)$ where $\psi$ is low-dimensional parameter vector.
   - Standard statistical theory is applicable.

Statistical Modelling of Spatio-Temporal Dependence

**Spatial Dependence**
- Assume route flows $x_1, \ldots, x_M$ are independent random variables.
  - In practice we may well see dependence, e.g. clustering.

**Temporal Dependence**
- Assume flows between days are independent (e.g. $x^s$ and $x^t$ are independent for $t \neq s$).
  - May expect autocorrelation in data.
  - No too difficult to incorporate this...
  - ... but many real data streams show remarkably little serial dependence.
Modelling Mean Routing Proportions

- $\mu = B\lambda$ can be inverted if we know mean routing proportions.
- $\lambda = P\mu$ where $P = (p_{ij}) = P(\theta)$ is matrix of mean routing proportions.
- Some authors assume $P$ known (or defined by some assignment model).
- We will assume $P(\theta)$ is unknown (a nuisance parameter).
  - Hence problem is estimation of $\lambda$ rather than $\mu$. 
Poisson Models

\[ x \sim \text{Pois}(\lambda) \quad (\text{interpret elementwise, with independence}) \]

Some properties

- \[ u \sim \text{Pois}(\mu) \]
- \[ x_i | u_j \sim \text{Bin}(p_{ij}, u_j) \text{ when } b_{ij} = 1 \]
- \[ \nabla(x) = \text{diag}(\lambda) \quad (\text{denotes diagonal matrix, } \lambda \text{ on diagonal}) \]
- \[ f(y) = \sum_{x: y = Ax} f(x) \quad (f \text{ denotes mass/density function}) \]

Some Comments

- Common choice for small counts.
- Strong assumption on mean-variance relationship.
- Link count distribution \( f(y) \) complicated.
Normal Models

\[ x \sim N(\lambda, \Sigma) \quad \text{(appropriately dimensioned multivariate normal)} \]

Some properties

- \( u \sim N(\mu, B\Sigma B^T) \)
- \( \nabla(x) = \Sigma \)
- \( y \sim N(A\lambda, A\Sigma A^T) \)

Some Comments

- Discretization error an issue for small counts.
- Flexibility in mean-variance relationship. E.g.
  - \( \Sigma \) assumed constant matrix (obtained how?)
  - \( \Sigma \) a function of \( \lambda \); e.g. Poisson approximation if \( \Sigma = \text{diag}(\lambda) \)
  - Estimate \( \Sigma \) too; current work by Shao Hu et al.
The Indeterminacy Problem

- Estimation simple if we directly observed \( \{x^t : t = 1, \ldots, n\} \).
  - E.g. define estimator \( \hat{\lambda} = \bar{x} = n^{-1} \sum_{t=1}^{n} x^t \)
- But only observe link counts \( \{y^t : t = 1, \ldots, n\} \) with
  \[
y^t = Ax^t \quad (t = 1, \ldots, n)
  \]
- \( A \) is \( N \) by \( M \) link-path incidence matrix.
  - Number links = \( N = \dim{(y)} \).
  - Number routes = \( M = \dim{(x)} \).
- Typically \( N << M \) so equations hugely underdetermined.
- Feasible route set \( \mathcal{X}_y = \{x: y = Ax\} \) may be massive.
A Toy Example

\[ y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{13} \\ x_{23} \end{bmatrix} = A \mathbf{x} \]
A Toy Example

\[ \begin{align*}
    y_1 &= 10 \\
    y_2 &= 10
\end{align*} \]

\[ y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{13} \\ x_{23} \end{bmatrix} = Ax \]

Conditional on \( y = (10, 10)^T \), latent route flows might be \( x = (0, 10, 0)^T \) or \( x = (10, 0, 10)^T \), or \( x = (5, 5, 5)^T \) etc.
What are the Consequences of Indeterminacy?

Some possible responses (for a general network)...

1. Link count data alone (e.g. without prior/target OD matrix) will never be able to provide a uniquely optimal estimator of $\lambda$.

2. It is possible to derive a unique estimator of $\lambda$ from a sufficiently long sequence of link counts $\{y^t: t = 1, \ldots, n\}$.

3. It is possible to derive a unique estimator of $\lambda$ from a single vector of link counts $y$.

Which of these is correct depends on modelling assumptions.
What are the Consequences of Indeterminacy?

Some possible responses (for a general network)...

1. Link count data alone (e.g. without prior/target OD matrix) will never be able to provide a uniquely optimal estimator of $\lambda$.
2. It is possible to derive a unique estimator of $\lambda$ from a sufficiently long sequence of link counts $\{y^t: t = 1, \ldots, n\}$.
3. It is possible to derive a unique estimator of $\lambda$ from a single vector of link counts $y$.

Which of these is correct depends on modelling assumptions.

For many reasonable models:

1. is false.
2. is true.
3. is sometimes true.
The Importance of second order properties

First Order Statistical Properties
- Mean count vector is $\bar{y} = n^{-1} \sum_{t=1}^{n} y^t$.
- Mean link counts provide just $N$ pieces of information.

Second Order Statistical Properties
- Sample variance is $S_y = A^T S_x A$
- $S_y$ provides $N(N + 1)/2$ pieces of information.
The Importance of second order properties

First Order Statistical Properties

- Mean count vector is $\bar{y} = n^{-1} \sum_{t=1}^{n} y^t$.
- Mean link counts provide just $N$ pieces of information.

Second Order Statistical Properties

- Sample variance is $S_y = A^T S_x A$
- $S_y$ provides $N(N + 1)/2$ pieces of information.
- $S_y$ can provide much additional information.
- Hence estimation from sequence $\{y^t: t = 1, \ldots, n\}$ plausible without prior information.


Correlation when $n = 1$?

Recall toy example:

Consider just two possibilities for route flow vector:

1. $\mathbf{x} = (10, 0, 10)^T$, i.e. 10 cars $1 \rightarrow 2$; 10 cars $2 \rightarrow 3$; 0 cars $1 \rightarrow 3$.

2. $\mathbf{x} = (0, 10, 0)^T$, i.e. 0 cars $1 \rightarrow 2$; 0 cars $2 \rightarrow 3$; 10 cars $1 \rightarrow 3$. 
Correlation when \( n = 1 \)?

Recall toy example:

\[
\begin{array}{c}
1 \\
y_1 = 10
\end{array} \quad \begin{array}{c}
1 \\
y_2 = 10
\end{array} \\
\begin{array}{c}
2 \\
1 \\
2 \\
3
\end{array}
\]

Consider just two possibilities for route flow vector:

1. \( x = (10, 0, 10)^T \), i.e. 10 cars 1 → 2; 10 cars 2 → 3; 0 cars 1 → 3.
2. \( x = (0, 10, 0)^T \), i.e. 0 cars 1 → 2; 0 cars 2 → 3; 10 cars 1 → 3.

Possibility 2 more intuitive because link counts are equal.

Maximum Likelihood Estimation

- Method requires no prior (target) OD matrix.
- Likelihood function is density/mass function of counts, viewed as function of parameters. E.g. assuming independence,

\[ L(\lambda) = \prod_{t=1}^{n} f(y^t|\lambda) \]

- Usually work with log-likelihood

\[ \ell(\lambda) = \log\{L(\lambda)\} = \sum_{t=1}^{n} \log\{f(y^t|\lambda)\} \]

- Maximum likelihood estimator (MLE) is \( \hat{\lambda} = \text{argmin}_\lambda L(\lambda) \).
- Optimal estimation under regular statistical theory.
- Measures of precision (e.g. confidence intervals) available.
Maximum Likelihood Estimation: Previous Work

- First big mention in transport literature by Spiess (1987)
  - But his approach confused parameters and random variables
- Landmark contribution by Yehuda Vardi (1996)
  - Preferred quasi method of moments approach $n$ small
- Hazelton (2000) extended Vardi’s work (to general routing)
- Studied for communication networks, e.g. Castro et al. (2004)


Maximum Likelihood Estimation: Summary

- For **normal model** with fixed covariance matrix $\Sigma$, unique maximum likelihood estimation not possible.

- For **Poisson model**, unique maximum likelihood theoretically possible for sufficiently large $n$.
  - Theory demonstrates statistically identifiability (Vardi, 1996; Hazelton, 2000).
  - Results disappointing when $n$ small.
  - Severe computational problems makes implementation impossible for large networks.

- Use of normal model with $\Sigma = \Sigma(\lambda)$ more promising, but sensitive to specification of mean-variance relationship.
Method of Moments

- If $\Sigma = \Sigma(\lambda)$ then simultaneously solve

$$\mathbb{E}[y] = A\lambda = \bar{y} \quad \text{and} \quad \var{y} = A\Sigma(\lambda)A^T = S_y$$


to get method-of-moments estimator.

- Typically those equations will not have a solution, so try minimizing function like

$$Q(\lambda) = \|A\lambda - \bar{y}\|^2 + \phi \|\text{vec}(A\Sigma(\lambda)A^T) - \text{vec}(S_y)\|^2$$

where $\phi$ is a weight and vec returns vector of matrix elements.


- Specification of $\Sigma(\lambda)$ is critical.

- Measures of precision for estimates?
Overview of Bayesian Inference

- In Bayesian paradigm, \( \lambda \) is a random vector.
  - Distribution of \( \lambda \) represents our knowledge/beliefs about it.
- Any existing knowledge expressed by prior, density \( f(\lambda) \).
  - Provides principled method for incorporating additional information to counter indeterminism problems.
- After observing link counts \( y \), distribution updates to posterior: \( f(\lambda|y) \).
- Posterior mode or mean can be used as point estimate of \( \lambda \).
- Posterior distribution can produce measures of precision (e.g. credible intervals).
- Being likelihood-based, Bayesian methods automatically incorporate second (and higher) order properties of the data.
The Posterior Distribution

- Posterior related to prior by

\[
f(\lambda | y) = \frac{f(y | \lambda) f(\lambda)}{f(y)}
\]

- Recall \( f(y | \lambda) = L(\lambda) \) is model likelihood.
- \( f(y) = \int f(y | \lambda) f(\lambda) d\lambda \) is just a normalizing constant.

- Bayesian approach introduced in transport literature by Maher (1983).

Bayesian Inference for Poisson Models

- For Poisson model $y \sim \text{Pois}(\lambda)$, cannot generally be computed.
- This is because it requires computation of likelihood

$$L(\lambda) = \sum_{y \in \mathcal{X}(y)} f(x | \psi)$$

and set $\mathcal{X}_y = \{x : y = Ax\}$ typically too big to enumerate.
MCMC Methods for Poisson Models

- Can proceed using Markov chain Monte Carlo (MCMC).
  - This is a methodology for sampling from posterior when posterior not available in closed form.
- Requires efficient methods for sampling feasible route flows (from $\lambda_y$).
- Difficult, but some progress from seminal paper by Tebaldi and West (1998), and also Hazelton (2010).


Bayesian Inference for Normal Models

- Model is $\mathbf{x} \sim \mathcal{N}(\mathbf{\lambda}, \Sigma)$.
- When $\Sigma$ is fixed (not dependent on $\mathbf{\lambda}$) posterior is available in closed form (Maher, 1983).
  - But with fixed $\Sigma$ we need more prior information for estimation, since correlation structure not informative.
- When $\Sigma = \Sigma(\mathbf{\lambda})$ then $f(\mathbf{y}) = \int f(\mathbf{y}|\mathbf{\lambda}) f(\mathbf{\lambda}) d\mathbf{\lambda}$ cannot be evaluated directly.
- Can use Markov chain Monte Carlo methods to sample from posterior.
  - But will not work efficiently unless we sample candidate values near to feasible route flow set.
Measurement Error

- Traffic counters not 100% reliable.
- Measurement error may render traffic counts inconsistent.

Possible Approaches

1. Pre-process link counts; e.g. Jörnsten and Stein (1993).
   - Artificially removes a source uncertainty, so can effect confidence intervals etc.

2. Incorporate measurement error in statistical model.
   - Fine for normal model: e.g. \( y \sim \mathcal{N}(A\lambda, A\Sigma A\lambda + I\sigma^2_\varepsilon) \) where \( \sigma^2_\varepsilon \) is measurement error variance (Hazelton, 2001).
   - Difficult with Poisson model.


Suppose we have some routing information from e.g. tracking GPS equipped vehicles.

Let $p$ be vector of probabilities of vehicle tracking for each route.

If exogenous estimates of $p > 0$ are available, then identifiability problems in theory addressed.

Where available, important to include both sporadic routing and link count data.

When collected contemporaneously, creates two part likelihood:

$$L(\lambda, p) = f(y_{not} | \lambda, p) \cdot f(x_{trk} | \lambda, p)$$
Example

- Observe data $y = (10, 10)^T$ and $x_{trk} = (1, 1, 1)$, so that $y_{not} = (8, 8)^T$.
- Use normal model with fixed covariance matrix $x \sim N(\mu, \Sigma)$
- Likelihood from link counts only will not have unique maximum in this case.
Example (Profile) Log-Likelihood Without Routing Information

Ridged – complete lack of identifiability.
Example (Profile) Log-Likelihood With Routing Information

Curvature introduced, and hence unique maximum likelihood estimate obtained.
Further Comments on Incorporation of Routing Information

- Problems much harder when $p$ is not known.
- In extreme case, routing information is then of no help at all!
- In practice we can get somewhere by using simple (perhaps crude) models for $p$.
- Analysis done in collaboration with Katharina Parry.

Acknowledgements

Support from the Royal Society of New Zealand (Marsden fund) gratefully acknowledged.

Katharina Parry (IIMS, Massey University, New Zealand)
References


References (continued)


For a copy of these slides...

http://www.massey.ac.nz/~mhazelto/seminars