Calibration of Dynamic Traffic Assignment Models

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

Presented at DADDY 09, Salerno Italy, 2-4 December 2009

Outline

Part 1: Statistical Inference as Part of the Modelling Process

Two Stages of Model Building

1. Development of mathematical description of process;
2. Calibration – i.e. estimation of unknown model parameters.

Two Stages of Model Building

1. Development of mathematical description of process;
2. Calibration – i.e. estimation of unknown model parameters.

Thoughts on the Above

- Transport research literature has tended to focus more heavily on 1 than 2.
- Both stages are equally important.

Methods of Calibrating Assignment Models

Stochastic Assignment Models

- Traffic flows modelled as random variables.
- Standard statistical methodologies can be applied (in theory).
  - Maximum likelihood estimation
  - Least squares estimation
  - Method of moments

Deterministic Assignment Models

- More difficult to apply principled methodology
- One approach is to embed in stochastic model and fit that
  - E.g. SUE as approximate mean of Markov assignment process.

Methods of Calibrating Assignment Models

Stochastic Assignment Models

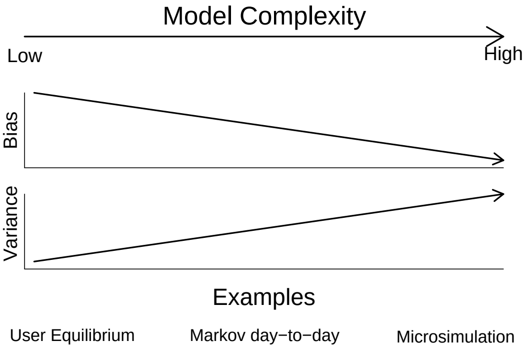
- Traffic flows modelled as random variables.
- Standard statistical methodologies can be applied (in theory).
  - Maximum likelihood estimation
  - Least squares estimation
  - Method of moments

Deterministic Assignment Models

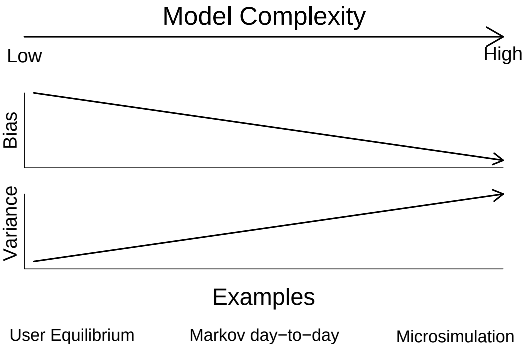
- More difficult to apply principled methodology
- One approach is to embed in stochastic model and fit that
  - E.g. SUE as approximate mean of Markov assignment process.

Assume henceforth that models are stochastic

Model Complexity



Model Complexity



Bias-Variance Trade-Off

$$MSE = bias^2 + var$$

The Dangers of Over-fitting

- Excessively complex models lead to over-fitting.
- Over-fitted models are deceptively 'realistic'.
- Excellent at reproducing yesterday.
- Poor at forecasting tomorrow.

The Dangers of Over-fitting

- Excessively complex models lead to over-fitting.
- Over-fitted models are deceptively ‘realistic’.
- Excellent at reproducing yesterday.
- Poor at forecasting tomorrow.

- So ... a Markov day-to-day model of traffic flows may be better in practice than a microsimulation.

Estimation, Reconstruction and Prediction

Aims in Model Fitting

- Estimation of model parameters with minimum error.
- Forecasting future realized flows.

Estimation, Reconstruction and Prediction

Aims in Model Fitting

- Estimation of model parameters with minimum error.
- Forecasting future realized flows.

Reconstruction of historical realized flows much less important.

Preparatory Notation

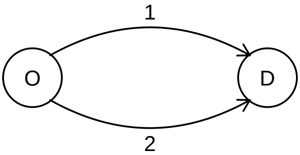
Random Variables

$\boldsymbol{u} = (u_1, \ldots, u_L)^\mathsf{T}$	OD flows
$\boldsymbol{y} = (y_1, \ldots, y_M)^\mathsf{T}$	route (path) flows
$\boldsymbol{x} = (x_1, \ldots, x_N)^\mathsf{T}$	link (arc) flows

Model Parameters

$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_L)^\mathsf{T}$	mean OD flows
$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)^\mathsf{T}$	mean route flows
$\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L$	route choice probability vectors by OD pair

An Illustrative Toy Example



Aim

- To model hourly traffic flow from O to D by paths 1 and 2.

Model Structure

- OD demand model: $u \sim \text{Pois}(\mu)$.
- Route choice: travellers take route 1 with prob. p_1.
- Hence $y_1|u \sim \text{Bin}(u, p_1)$.

Toy Example: Model Parameters

Known Parameter

- $\mu = 10$ travellers per hour.

Unknown Parameters (need estimating)

- Route choice prob. p_1^t varies from hour to hour.

Available Data

- Hourly counts: $\boldsymbol{y}^t = (y_1^t, y_2^t)^\mathsf{T}$ for $t = 1, 2, \ldots, 24$ hours.

Toy Example: Models to be Calibrated

Hour-to-Hour Model

- Model correctly specified.
- Model complex: 24 unknown parameters to be estimated.

Day-to-Day Model

- Just model aggregate path flows over whole day.
- Assume that $p_1^t = p_1$, a constant (counterfactual).
- Model mis-specified.
- Model simple: just 1 parameter to be estimated.

Toy Example: Fitted Model Comparison

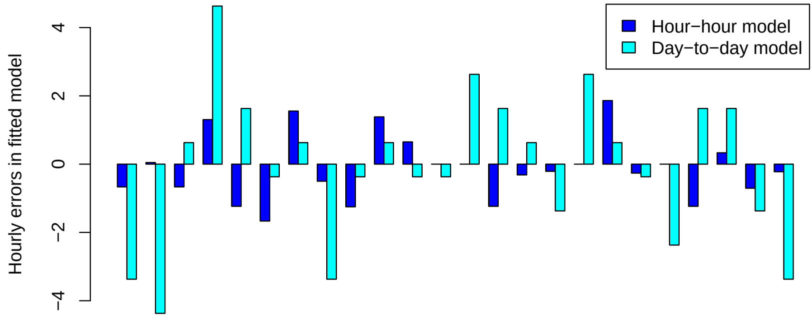
Parameter Estimation

- Estimate route choice probability by maximum likelihood: $\hat{p}_1^t$. For day-to-day model, $\hat{p}_1^t = \hat{p}_1$.

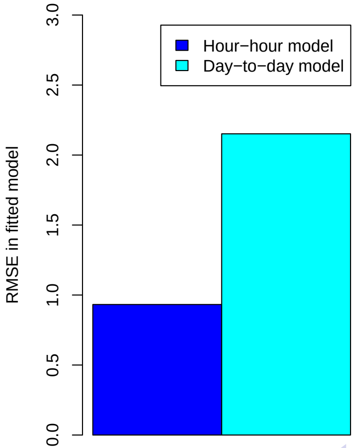
Fitted Model Comparison

- Today's reconstruction errors: $y_1^t - \mu\hat{p}_1^t$
- Tomorrow's predictive errors: $\breve{y}_1^t - \mu\hat{p}_1^t$
- Week average predictive errors: $\bar{y}_1^t - \mu\hat{p}_1^t$

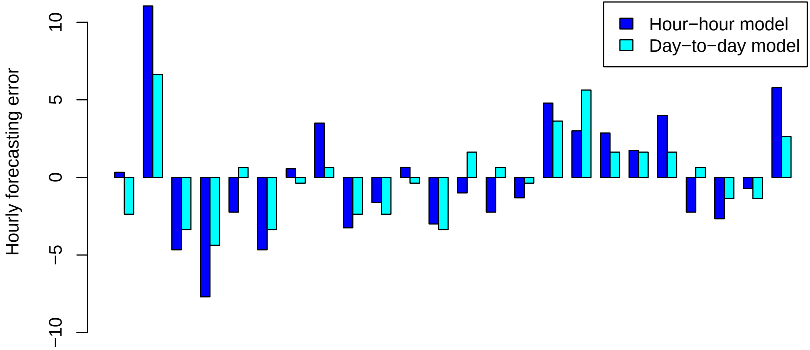
Toy Example: Hourly Errors in Reconstruction



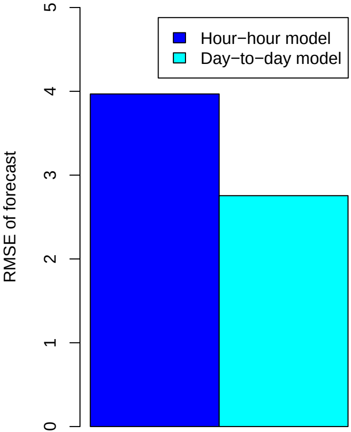
Toy Example: Aggregate Error in Reconstruction



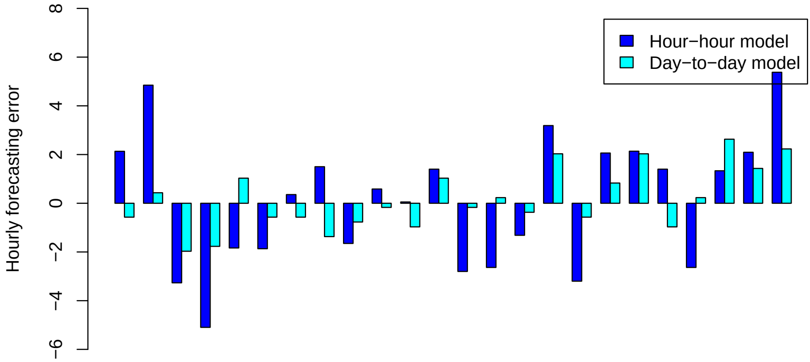
Toy Example: Hourly Errors for Tomorrow



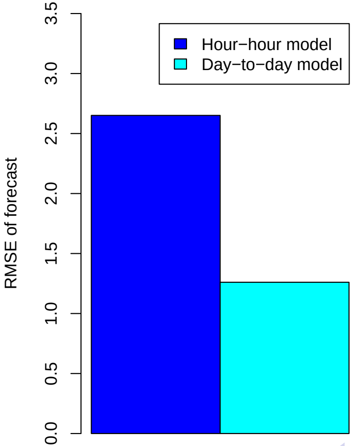
Toy Example: Aggregate Error for Tomorrow



Toy Example: Hourly Errors for Next Week



Toy Example: Aggregate Error for Next Week



Toy Example: Summary of Results

- Complex (hour-to-hour) model is great at forecasting yesterday.
- Simple (day-to-day) model is much better at predicting tomorrow.

A General Conclusion

Model design should account for feasibility of good calibration.

Part 2: Some General Thoughts on Inference

Some General Thoughts on Inference

- Data sources
- Model parameterization
- Link counts and indeterminism
- The Importance of second order properties
- Linear inverse framework

Data

- **Link count data**
  - Widely available
  - Typically unbiased
- **Vehicle routing information**
  - Availablility varies
  - Can be biased
- **Other**
  - Surveys (bias? coverage?)
  - Experiments

Data

- **Link count data**
  - Widely available
  - Typically unbiased
- **Vehicle routing information**
  - Availablility varies
  - Can be biased
- **Other**
  - Surveys (bias? coverage?)
  - Experiments

We will focus primarily on inference from link count data.

Model Parameterization

- Some parameters can be estimated directly from link counts
  - E.g. cost (delay) functions
- Behavioural parameters control route choice, hence route counts provide direct information

Model Parameterization

- Some parameters can be estimated directly from link counts
  - E.g. cost (delay) functions
- Behavioural parameters control route choice, hence route counts provide direct information

Example (logit route choice)

$$p_i = \frac{\exp(-\theta c_i)}{\sum_j \exp(-\theta c_j)}$$

Parameter θ a behavioural parameter.

Link counts and indeterminism

Fundamental equation

$$\boldsymbol{x} = A\boldsymbol{y}$$

- $A = (a_{ij})$ is routing matrix.
  - $a_{ij} = 1$ if link i on route j, 0 otherwise.
- Number links $= N = \dim(\boldsymbol{x})$.
- Number routes $= M = \dim(\boldsymbol{y})$.
- Typically $N << M$ so equations hugely underdetermined.
- *Feasible route set* $\mathcal{Y}_{\boldsymbol{x}} = \{\boldsymbol{y} : \boldsymbol{x} = \mathcal{A}\boldsymbol{y}\}$ can defy enumeration.

The Importance of second order properties

Data

$\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^n$ sequence of link counts

First Order Statistical Properties

$$\bar{\boldsymbol{x}} = A\bar{\boldsymbol{y}}$$

- Mean link counts provide just N pieces of information.

The Importance of second order properties

Second Order Statistical Properties

$$S_{\boldsymbol{x}} = A^T S_{\boldsymbol{y}} A$$

- Sample variance provides $N(N-1)/2$ pieces of information.

The Importance of second order properties

Second Order Statistical Properties

$$S_{\boldsymbol{x}} = A^T S_{\boldsymbol{y}} A$$

- Sample variance provides $N(N-1)/2$ pieces of information.

Conclusion

Second order properties provide lots of additional information.

Linear Inverse Framework

Statistical Linear Inverse Problem

$$q(\boldsymbol{x}) = \int h(\boldsymbol{x}, \boldsymbol{y})\, dP(\boldsymbol{y})$$

- P is probability measure for latent variables $\boldsymbol{y}$
- h is *blurring* function
- q is density/mass function for observed variables $\boldsymbol{x}$.
- Examples:
  - Image deblurring
  - Decomposition of chemical spectra

Linear Inverse Problems in Transport

$$q(\boldsymbol{x}) = \int h(\boldsymbol{x}, \boldsymbol{y})\, dP(\boldsymbol{y})$$

- $P(\boldsymbol{y})$ probability measure for route flows
  - possibly over multiple days
- $q(\boldsymbol{x})$ probability density/mass function for link flows.
- $h(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{1}_{\boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{x}}}$ for error-free counts.
- E.g. $h(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x} - A\boldsymbol{y})$ for counts with measurement error.

Statistical Linear Inverse Problems (SLIPs)

- Puts inference for transport networks in wider context.
- Lots known about these problems ...
  - SLIPs are hard
  - Regularization typically necessary
  - Bayesian framework attractive
  - Each problem is different
- ... but much remains to be done.

Part 3: Inference with Small Traffic Counts

A Day-to-Day Assignment Model

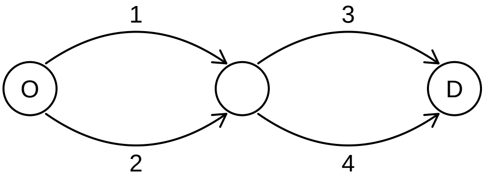
Markov Process Model

- Assume traffic pattern evolves as Markov process from day-to-day.
- Route (link) flows on day t are $\boldsymbol{y}^t$ ($\boldsymbol{x}^t$).
- Route travel costs experienced on day t are $\boldsymbol{c}^t = \boldsymbol{c}^t(\boldsymbol{x}^t)$.

Transition Probabilities

- Probability distribution of $\boldsymbol{y}^t$ specified in term of
  - ν previous travel costs: $\boldsymbol{c}^{t-1}, \ldots, \boldsymbol{c}^{t-\nu}$;
  - Parameter vector $\boldsymbol{\theta}$, requiring estimation.
- Denote by $p(\boldsymbol{y}^t | \boldsymbol{c}^{t-1}, \ldots, \boldsymbol{c}^{t-\nu}, \boldsymbol{\theta})$.

Figure-of-Eight Example



Route	Constituent links	Route cost
1	1,3	$c_1 = k_1 + k_3$
2	1,4	$c_2 = k_1 + k_4$
3	2,3	$c_3 = k_2 + k_3$
4	2,4	$c_4 = k_2 + k_4$

Figure-of-Eight Example

OD Demand $u^t \sim \text{Pois}(\mu)$

Figure-of-Eight Example

OD Demand $u^t \sim \text{Pois}(\mu)$

Route Choice – depends on yesterday's costs (first order Markov)

$$p_i^t \equiv p_i^t(\zeta) = \frac{e^{-\zeta c_i^{t-1}}}{\sum_{j=1}^{N} e^{-\zeta c_j^{t-1}}}$$

Figure-of-Eight Example

OD Demand $u^t \sim \text{Pois}(\mu)$

Route Choice – depends on yesterday's costs (first order Markov)

$$p_i^t \equiv p_i^t(\zeta) = \frac{e^{-\zeta c_i^{t-1}}}{\sum_{j=1}^{N} e^{-\zeta c_j^{t-1}}}$$

Parameter vector $\boldsymbol{\theta} = (\mu, \zeta)^\mathsf{T}$

Likelihood Based Inference

Likelihood

$$L(\boldsymbol{\theta}) = f(\boldsymbol{X}|\boldsymbol{\theta})$$

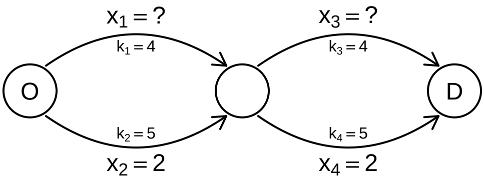
- f generically denotes probability mass/density function
- $\boldsymbol{X} = (\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n)$ is all link data.
- Parameter $\boldsymbol{\theta}$ describes *route* flows.

Decomposition

$$f(\boldsymbol{X}|\boldsymbol{\theta}) = \sum_{\boldsymbol{Y}} f(\boldsymbol{X}|\boldsymbol{Y}) f(\boldsymbol{Y}|\boldsymbol{\theta}) = \sum_{\boldsymbol{Y} \in \mathcal{Y}_{\boldsymbol{X}}} f(\boldsymbol{Y}|\boldsymbol{\theta})$$

- $\mathcal{Y}_{\boldsymbol{X}} = \{\boldsymbol{x}^t = A\boldsymbol{y}^t : t = 1, \ldots, n\}$ is feasible route set.

Application to Figure-of-Eight Example



Simple example (for clarity)

- Link count data from just one day.
- Counts available on links 2 and 4 only: $\boldsymbol{x} = (\mathtt{NA}, 2, \mathtt{NA}, 2)^\mathsf{T}$.
- Link costs fixed: $\boldsymbol{k} = (4, 5, 4, 5)^\mathsf{T}$
- Is this sufficient information to estimate μ and ζ?

Likelihood for Figure-of-Eight Example

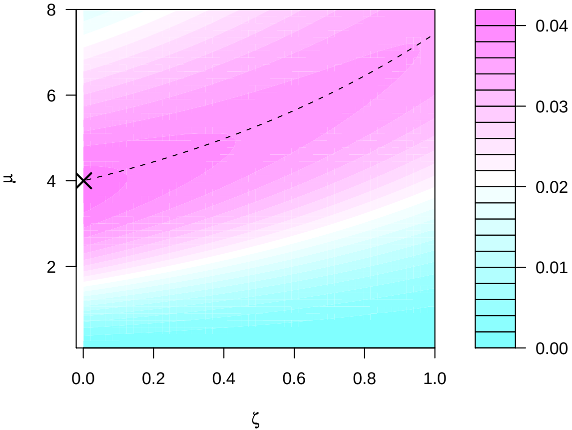
$$L(\mu, \zeta) = \sum_{\boldsymbol{y} \in \mathcal{Y}} \prod_{i=1}^{4} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

- Feasible set $\mathcal{Y} = \{(y_1, y_2, y_3, y_4)^\mathsf{T} : y_2 + y_4 = 2, y_3 + y_4 = 2\}$.
- Can sum out unobserved y_1.
- Then $\mathcal{Y} = \{\boldsymbol{y} : y_2 = 2 - y_4, y_3 = 2 - y_4, y_4 = 0, 1, 2\}$

$$L(\mu, \zeta) = \sum_{y_4 \in \{2,3,4\}} \frac{e^{-(\mu_2+\mu_3+\mu_4)} \mu_2^{2-y_4} \mu_3^{2-y_4} \mu_4^{y_4}}{(2-y_4)!(2-y_4)!y_4!}$$

Normalized Likelihood for Example

- Dashed line is set of GLS estimates.
- Likelihood has unique maximum at $(\mu, \zeta) = (4, 0)$.



Computational Problems

- Likelihood based inference desirable (see example).
- Evaluation of full likelihood requires enumeration of all feasible routes.
- Only feasible for very small examples.
- In general, direct likelihood approach is impractical.

Bayesian Approach

- In Bayesian paradigm, parameters are random variables.
- Distribution of parameter represents current knowledge about it.
- Before data collected, knowledge given by *prior distribution* $f(\boldsymbol{\theta})$.
- After data $\boldsymbol{X}$ observed, knowledge given by posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{X})$.

Calculating the Bayesian Posterior

$$f(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{X})} \propto L(\boldsymbol{\theta})f(\boldsymbol{\theta})$$

Bayesian MCMC

- Bayesian inference cannot proceed directly without likelihood $L(\boldsymbol{\theta})$.
- Computationally feasible alternative is to sample from posterior.
- Can do this using Markov chain Monte Carlo (MCMC) methods.

Implementing MCMC

- Must jointly sample parameters $\boldsymbol{\theta}$ and route flows $\boldsymbol{Y}$ conditional on $\boldsymbol{X}$.
- Sampling $\boldsymbol{Y}$ given $\boldsymbol{X}$ is challenging since $\mathcal{Y}_{\boldsymbol{X}}$ not enumerable.
- Working in progress.
- See presentation by Katharina Parry.

Part 4: Large Count Approximations

Normal Approximations

$$\begin{aligned} \boldsymbol{u} &\sim \text{Pois}(\mu) \\ \boldsymbol{y}_i | u_i &\sim \text{Mult}(\boldsymbol{p}_i) \\ \Rightarrow \boldsymbol{y}_i &\sim \text{Pois}(\mu_i \boldsymbol{p}_i) \end{aligned}$$

Define: $\boldsymbol{\lambda}_i = \mu_i \boldsymbol{p}_i, \quad \boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\mathsf{T}, \ldots, \boldsymbol{\lambda}_M^\mathsf{T})^\mathsf{T}$.

Then approximately for large $\boldsymbol{\lambda}$:

$$\boldsymbol{y} \,\dot{\sim}\, \text{N}\left(\boldsymbol{\lambda}, \text{diag}(\boldsymbol{\lambda})\right)$$

Normal Approximation Magic?

Large counts

$$
\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{\theta}) &\approx \mathsf{N}\left(\boldsymbol{\lambda}, \mathsf{diag}(\boldsymbol{\lambda})\right) \\
\Rightarrow f(\boldsymbol{x}|\boldsymbol{\theta}) &\approx \mathsf{N}\left(A\boldsymbol{\lambda}, A\mathsf{diag}(\boldsymbol{\lambda})A^{\mathsf{T}}\right)
\end{aligned}
$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta})$.

Small counts

$$
f(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{x}}} f(\boldsymbol{y}|\boldsymbol{\theta})
$$

Normal Approximation Magic?

Large counts

$$
\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{\theta}) &\approx \mathsf{N}\left(\boldsymbol{\lambda}, \mathrm{diag}(\boldsymbol{\lambda})\right) \\
\Rightarrow f(\boldsymbol{x}|\boldsymbol{\theta}) &\approx \mathsf{N}\left(A\boldsymbol{\lambda}, A\mathrm{diag}(\boldsymbol{\lambda})A^{\mathsf{T}}\right)
\end{aligned}
$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta})$.

Small counts

$$
f(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{x}}} f(\boldsymbol{y}|\boldsymbol{\theta})
$$

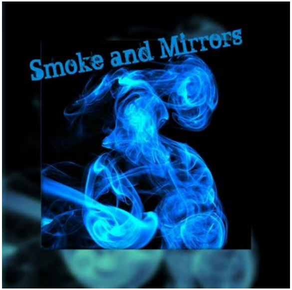
- Large sample distribution looks much more tractable.
- Is this magic??

Smoke and Mirrors

Smoke and Mirrors

The apparent advantage of the large count likelihood is partly an illusion.

- Even if link counts large, what about path flows?
- Complexity is hidden in mean $A\boldsymbol{\lambda}$ and covariance matrix $A\text{diag}(\boldsymbol{\lambda})A^{\mathsf{T}}$.



Application to Figure-of-Eight Example

$$\ell(\mu,\zeta) = \log(L(\mu,\zeta)) = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{m}) + \text{const.}$$

where

$$\boldsymbol{x} = \begin{bmatrix} x_2 \\ x_4 \end{bmatrix} \qquad \boldsymbol{m} = \mu\begin{bmatrix} p_3+p_4 \\ p_2+p_4 \end{bmatrix} \qquad \Sigma = \mu\begin{bmatrix} p_3+p_4 & p_4 \\ p_4 & p_2+p_4 \end{bmatrix}$$

and

$$p_i(\zeta) = \frac{e^{-\zeta c_i}}{\sum_{j=1}^{N} e^{-\zeta c_j}}$$

Lessons from the Example

- Likelihood is a complex function of μ, ζ even in simple example.
- Even if mean and variance estimated well, may be difficult to draw conclusions about canonical parameters.

Normal Approximations for Day-to-Day Assignment

Theorem (from Davis and Nihan (1993)[1])

For fixed demand $\boldsymbol{\mu}$, Markov assignment process $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots$ can be approximated by a normal vector autoregressive (VAR) process.

In other words:

$$\boldsymbol{x}^t | \boldsymbol{x}^{t-1}, \ldots, \boldsymbol{x}^{t-\nu} \dot{\sim} \mathsf{N}(\boldsymbol{m}^t, \Sigma^t)$$

where $\boldsymbol{m}^t$, Σ^t functions of $\boldsymbol{\theta}$ and $\boldsymbol{x}^{t-1}, \ldots, \boldsymbol{x}^{t-\nu}$.

[1] Davis G. and Nihan N. (1993). *Op Res* **41** 169–178.

Inference Using VAR Approximation

- Earlier comments notwithstanding, VAR approximation provides best current hope for inference for day-to-day models.
- Computation of VAR process mean vector and covariance matrix is challenging.
- Dealing with terms without full history $(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^\nu)$ difficult.
- Does VAR approximation work for Poisson (etc.) demand?

Part 5: Conclusions and Future Directions

Conclusions

- Parameter estimation is a critical step in modelling day-to-day traffic patterns.
- Statistics inference is challenging.
- Problems inevitable – dealing with large scale linear-inverse problems.
- Methods for small counts and large counts differ markedly.
- There remain many more questions than answers.

Future Directions

- MCMC seems best hope for inference for small count models.
- Good sampler for route flows is crucial.
- Try VAR approximation for large flows.
- Need better understanding of VAR model properties.

Acknowledgement

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

For a copy of these slides...

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