Statistical Methods in Transportation Research

Martin Hazelton

martin@maths.uwa.edu.au
M.Hazelton@massey.ac.nz
The Travelling Statistician Problem

Perth

Palmerston North
In developed countries, transportation typically accounts between 5% and 12% of GDP.

Transport accounts for around 20% of global CO\textsubscript{2} emissions.

Around 2000 Australian transport fatalities each year (90% road).

Transport sector supplied 427000 jobs in Australia in 2003.

In 2000, road traffic congestion in U.S. estimated to have caused:

- 3.6 billion vehicle-hours of delay
- 21.6 billion litres of wasted fuel
- $67.5 billion in lost productivity

Statisticians account for 0.1% of transportation research. (OK, that one’s a guess!)
Transportation research (and what’s currently wrong with it . . .)

The audience buys a car

Some problems for statisticians:

- Speed estimation
- Estimation of trip matrices
- Modelling day-to-day dynamics of traffic flow

Further statistical problems
Variety of research fields:
- Transport policy
- Psychology and transport
- Transport and urban planning
- Transport modelling
Variety of research fields:
- Transport policy
- Psychology and transport
- Transport and urban planning
- Transport modelling

Most researchers transport modelling have an engineering or applied mathematics background.

Variability frequently gets ignored.
Examples of (stochastic) variation:

- Traffic counts
- Vehicle speeds
- Route choice
Examples of (stochastic) variation:

- Traffic counts
- Vehicle speeds
- Route choice

Ignore variation and:

- Extreme events may be missed (e.g. gridlock)
- Estimates and predictions may be biased
- Essential information may be lost
Variability is a statistician’s trade

Statisticians have the opportunity to play an important role in transportation research
The Audience Buys a Car

- **Aim:** to get best fuel efficiency

- **Dealer 1:**
  - Buy car that does 40 km per litre with prob. $\frac{1}{2}$
  - Buy car that does 10 km per litre with prob. $\frac{1}{2}$

- **Dealer 2:**
  - Buy car that does 20 km per litre
The mean km per litre are:

- **Dealer 1**: \( \frac{1}{2} \times 40 + \frac{1}{2} \times 10 = 25 \).
- **Dealer 2**: 20.

So get more km per litre on average going to Dealer 1.
Dealer 1:
- Buy car that needs 1/40 litre per km with prob. 1/2
- Buy car that needs 1/10 litre per km with prob. 1/2

Dealer 2:
- Buy car that needs 1/20 litre per km

The mean litres per km are:
- Dealer 1: \( \frac{1}{2} \times \frac{1}{40} + \frac{1}{2} \times \frac{1}{10} = \frac{1}{16}. \)
- Dealer 2: \( \frac{1}{20}. \)

So need less litres per km on average from Dealer 2.
Answers are different because generally

$$\frac{1}{\mathbb{E}[X]} \neq \mathbb{E}[1/X]$$

for a random variable $X$.

Car buying example illustrates importance of accounting for variation.
Automated vehicle detector set in road.

Detector is ‘occupied’ while vehicle passes over it.

Aggregate data returned for 20 or 30 second intervals:

- Vehicle count during interval, \( n \)
- Total time that detector is occupied during interval, \( y \).

**Goal**: to estimate mean vehicle speed during interval.
Each vehicle occupies detector for time it takes to travel the ‘effective vehicle length’ $\lambda$.

‘Typical’ effective vehicle length given by

$$\lambda = d + \bar{l}$$

where $d$ is detector range, $\bar{l}$ is mean vehicle length.
Speed Estimation: Simple Solution

\[
\text{Speed} = \frac{\text{Distance}}{\text{Time}}
\]

Hence simple estimator of speed is

\[
\hat{s} = \frac{n\lambda}{\bar{y}} = \frac{\lambda}{\bar{y}}
\]

where \( \bar{y} \) is mean occupancy of each vehicle.
Speed Estimation: Simple Solution

\[ \text{Speed} = \frac{\text{Distance}}{\text{Time}} \]

Hence simple estimator of speed is

\[ \hat{s} = \frac{n\lambda}{y} = \frac{\lambda}{\bar{y}} \]

where \( \bar{y} \) is mean occupancy of each vehicle. But note

\[ \hat{s} = \frac{\lambda}{\bar{y}} \neq \frac{[\lambda]}{[y]} = \bar{s} \]
Simple solution gives noisy estimation through time

Fails to properly account for variation in vehicle lengths.
A statistician might:

- Try and apply smoothing through time.
- Model individual vehicle speeds and lengths as missing data.
- Not so hard to do post hoc, perhaps with MCMC for model fitting.
- But traffic control often needs real-time estimates.
A statistician might:

- Try and apply smoothing through time.
- Model individual vehicle speeds and lengths as missing data.
- Not so hard to do post hoc, perhaps with MCMC for model fitting.
- But traffic control often needs real-time estimates.

Research problem: develop real-time speed estimation methods that are computationally cheap.
Trip Matrix Estimation: The Problem

- Estimate intensity of traffic flow between origins and destinations of travel on a network.
- Available data are traffic counts on certain road links over given time intervals.
Trip matrix easy to estimate from route traffic counts.
But route flows generally not uniquely determined from a single set of link counts.

Hence good trip matrix estimation impossible from aggregate link counts unless other information is available (e.g. informative prior).
Sequence of set of link counts provides vital additional information.

Trip matrix parameters are identifiable for e.g. Poisson model of trip generation.
What is a good stochastic model for trip generation?

Well founded statistical approaches largely restricted to estimation of static trip matrices.

Traffic control can require dynamic updating of trip matrices.
Trip Matrix Estimation: Challenges

- What is a good stochastic model for trip generation?
- Well founded statistical approaches largely restricted to estimation of static trip matrices.
- Traffic control can require dynamic updating of trip matrices.

Aim: to model time series of day-to-day traffic counts on network.

Model can be:

- Descriptive: used to understand properties of traffic system.
- Predictive

Example applications:

- estimate flow patterns after road closure;
- assess effect of proposed bypass.
Elements of Day-to-Day Model

- **Travel demand model**
  - trip matrix;
  - may be adjusted for seasonality, day of week etc.

- **Traffic assignment model**
  - describes how given demand will distributed itself over the network;
  - based on aggregation of route choices for all travellers.
Modelling Traveller Route Choice

- Route choice modelling pivotal.
- Natural to define model at microscopic (individual traveller) level.
- Travellers will select ‘cheap’ routes:
  - minimum travel time
  - short distance
  - easy drive
- (Random) variation in perceptions of costs between travellers.
- Properties of model required at macroscopic level (aggregate traffic flows).
Travellers base route choice on combination of travel costs experienced over finite history.

Inertia may suggest behaviour change only in case of clearly superior alternative.

**Pro:** model properties relatively well understood.

**Con:** no attempt to represent contemporaneous traveller interaction.
Parameters can represent:

- sensitivity to cost differences
- length of memory
- weighting of historical costs

Inference may be based on:

- traffic counts (very indirect information)
- stated preference surveys (believable?)
- travel diaries

Inference is far from straightforward in all cases.
Route flow evolve as Markov process.

Observed link flows linear combination of route flows, plus measurement error.

Hidden Markov model technology applicable, but:

- Model has peculiar structure;
- Practical problems may have huge size (e.g. thousands of road links, even more routes, for urban network model).
Route flow evolve as Markov process.

Observed link flows linear combination of route flows, plus measurement error.

Hidden Markov model technology applicable, but:

- Model has peculiar structure;
- Practical problems may have huge size (e.g. thousands of road links, even more routes, for urban network model).

Research problem: develop methodology for inference for route choice models from readily available data.
Applications include:

- number plate identification;
- automatic evaluation of road surface cracks.
Spatial statistics applied to:

- modelling road vehicle emissions in space and time;
- modelling health impacts (both globally and locally);
- modelling impact of road accidents with hazardous substances.
Further Statistical Problems in Transportation

- Survey design and analysis
- Random utility modelling and inference
- Stochastic operations research and system optimization
- Road accident modelling
- Transportation research likely to receive increasing attention as fuel prices and environmental costs spiral.
- Transportation science traditionally dominated by deterministic models and methods.
- Lots of opportunities for statisticians to make major contributions.