Graphical modelling for accident and exposure data

Creating the future of transport

Louise Walter, Affiliated to the University of Southampton

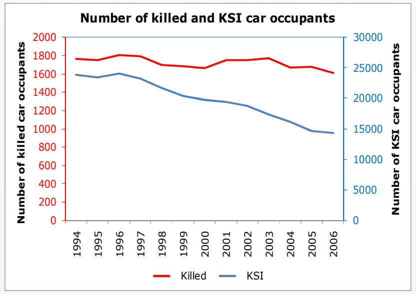
Motivation

Since the mid 1990s, the annual number of car occupants killed (K) in road accidents has remained fairly constant whilst the number of casualties being seriously injured (SI) continues to decrease.

A number of factors have been identified as contributing to this trend:

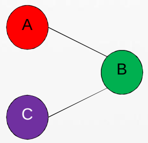
- single vehicle accidents
- 4x4 and people carrier occupants
- accidents at bends
- overturned cars

We investigate the causal effects of the slow decreasing fatal trend in Great Britain.



What is Graphical Modelling?

Graphical Modelling is a statistical technique which combines probability theory and graph theory, providing a way to build complex models by combining simpler parts. Data and prior knowledge are combined in Bayesian techniques to produce a network which joins 'linked' variables in the database with an edge in the graph. Knowing about these links helps to define which variables are important influencing factors.



Three variables A, B and C in a database may be represented by the Graphical Model shown.

The model says that A is conditionally independent of C given B: that is, if you know B, then knowing what A is will not help to predict C.

This technique is new to accident data. It was originally applied to gene network data and used for medical diagnostics but is now used in many areas where causal links between variables are of interest.

Graphical Modelling with STATS19

The STATS19 database is multi-dimensional and contains a considerable number of variables which are interrelated. The majority of current analyses are uni-variate, taking one or two variables at a time. These techniques are unable to deal with correlated information and are, consequently, simplistic.

The STATS19 database is an ideal candidate for Graphical Modelling due to these interrelated variables.

Models of accident data on their own have limited scope for interpretation, so the exposure data that are available (for example, vehicle kilometres) will be modelled as a separate Graphical Model, and then the two separate graphs will be linked together.

AIM
Model hierarchical accident data, taking into account available exposure data, by combining a series of graphical models, to assess which factors are influencing the fatal trends.

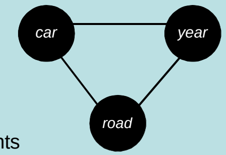
Step 1: Model exposure data

The exposure data consists of

- number of registered vehicles by car type (*car*) per year
- number of vehicle kilometres by road type (*road*) per year

A Monte Carlo Markov chain simulation technique estimates the number of vehicle kilometres by road type *and* car type per year.

Preliminary results suggest that the graphical model for exposure data links all three variables together.



Interpretation

Road – Car
Different types of car travel different amounts on different types of road.

Car – Year
The distribution of car types on the road network changes by year.

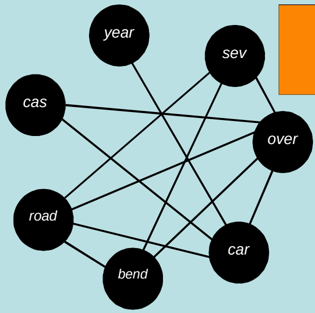
Year – Road
The number of vehicle kilometres travelled on different road types varies by year.

Step 2: Model accident data

The accident data consist of the variables year of accident (*year*), accident severity (*sev*), occupant type (*cas*) and road type (*road*) in addition to variables relating to the factors that were specified to be influencing the fatal trend:

- car type (*car*);
- accident at bend? (*bend*);
- car overturned? (*over*);

The early results show a highly linked graphical model which suggests many interdependencies in the dataset.



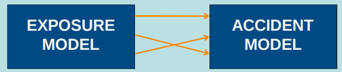
Interpretation (part)

Car – Year
The pattern of car type in accident changes by year (due to exposure?).

Also of interest are the missing links, for example, severity is not directly linked to (i.e. is conditionally independent of) the type of car or where you are sitting in that car.

Step 3: Join accident and exposure models together

Combining the models found in stages one and two, using a chain graph, will help to determine, for example, whether the changes in car type by year in the accident model match the dependence in the exposure data, taking all other variables into consideration.



FURTHER WORK

Currently Graphical Modelling techniques require the items in the data to be independent. This assumption only allows current techniques to be applied to single vehicle accidents.
Once the model, described above, has been finalised, it will be necessary to relax this independence assumption, and allow all casualties and collision types to be included in the model.