



Statistical Society of Australia Inc.

Victoria Branch Newsletter: November 2006

Contents

- 1) December Meeting
- 2) Bayesian Methods Workshop
- 3) September Meeting Report
- 4) Belz Lecture Report
- 5) Website

1) December meeting

This special Christmas meeting will be held at 6:15 p.m. on Tuesday 12 December 2006. It is a Joint meeting of the Victorian Branch of the SSAI with the Centre for Molecular, Environmental, Genetic and Analytic (MEGA) Epidemiology in the School of Population Health, University of Melbourne. It will be held in room 515, Level 5, 207 Bouverie Street (off Grattan Street one street west of Swanston Street), Carlton. The building is reference O18 on the campus map, which is available at <http://www.pb.unimelb.edu.au/CampusMaps/Parkville.pdf>

The meeting will be preceded by light refreshments and Christmas cheer at 5:45 p.m. Members and guests are invited to join the speaker for dinner at a nearby restaurant after the meeting.

Further enquiries can be directed to Lyle Gurrin on 8344 0731 or email lgurrin@unimelb.edu.au

Speaker: Bendix Carstensen, Steno Diabetes Center, Copenhagen, Denmark, bx@steno.dk
www.biostat.ku.dk/~bx

Title: Who Needs the Cox Model Anyway?

Abstract

In the last 30 years, survival analysis has been virtually synonymous with application of the Cox proportional hazards regression model. The common view of survival analysis (and teaching of it) from the Kaplan-Meier estimator to the Cox model is based on time as the response variable, incompletely observed (right-censored). This has automatically lent a certain aura of complexity to concepts such as time-dependent covariates, stratified analysis, delayed entry and time-varying coefficients.

If survival studies are viewed in light of the demographic tradition, the basic observation is not one of time to event (or censoring), but rather many small pieces of follow up from each individual. This makes concepts clearer, as modelling of rates rather than time to response becomes the focus; the basic response is now two dimensional: a 0/1 outcome in each interval and an interval length. Time is then correctly viewed as a covariate rather than a response. Time-dependent covariates will not have any special status relative to other covariates, except to the extent they are intermediate responses on which one conditions. Stratified analysis becomes a matter of interaction between time and a categorical covariate, and time-varying coefficients become interactions between time and a continuous covariate. Finally, the modelling tools needed reduces to Poisson regression (and ultimately logistic regression).

The Cox model may be viewed as a special case of a Poisson model where the detail in modelling of the time covariate has been taken ad absurdum, namely with one parameter per failure time. The main advantage of the demographic view is that researchers are forced to consider explicitly which timescale(s) to use and to what degree of detail it is relevant to model its effect on rates.

Contrary to this, Poisson modelling of disease rates and follow-up studies in epidemiology has traditionally (and 20 years ago for good computational reasons) been restricted to analysis of tables where rates have been assumed constant over fairly broad time spans, typically 5 years, as most methods have been developed in cancer epidemiology where 5 years is considered a short age-span. This approach is essentially one where initial tabulation of data unnecessarily limits the flexibility of modelling.

The practicalities of this change in modelling focus lead to a splitting of the follow-up period into many observations, and hence Poisson modelling of datasets with 10-50 times as many observations as persons. The lack of computing power has been a problem until recently, as has the absence of software. Through the teaching and software development efforts of David Clayton and Michael Hills, there are now solutions available for at least Stata, SAS and R, making this approach widely accessible.

The only remaining advantage of the Cox model is the ability to easily produce estimates of survival probabilities in (clinical) studies with a well-defined common entry time for all individuals. This can however also be produced from a Poisson model with little extra effort.

2) Bayesian Methods Workshop

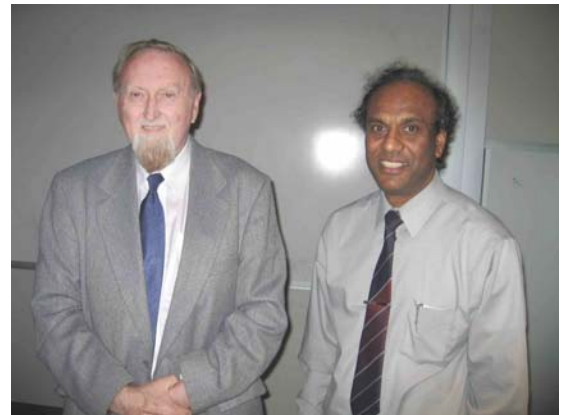
The VIC Branch is running a Bayesian Methods workshop on 28 and 29 November, 2006, 9.30 am-5.00 pm at the Swinburne University of Technology, Hawthorn Campus, Victoria. The presenter is Kerrie Mengersen, from Queensland University of Technology .

Places are limited due to the need for each participant to have use of a computer, so book ASAP. All details and the booking form are available at <http://www.statsoc.org.au/Branches/VIC/>

3) September Meeting Report

The 2003 Nobel Laureate in Economics, Professor Sir Clive Granger, was the special invited speaker at the meeting held on September 12 at Monash University Caulfield Campus. This event was jointly organized by the Department of Econometrics and Business Statistics at Monash University and the Statistical Society of Australia Victorian Branch. The presentation was attended by an enthusiastic audience of nearly a hundred.

The topic of the presentation was **"Building Policy Models for the Amazon Rain Forest"**. Professor Granger started the presentation by first noting that the Amazon rain forest is the largest in the world and is of considerable importance to global warming issues and is of economic importance to Brazil; however, it is being deforested at a rapid rate. Direct causes of deforestation are logging and illegal settlers and farmers.



Professor Granger (left) with Mervyn Silvapulle

The amount of money involved in deforestation is quite large indeed. According to *New York Times* (October 16, 2005), the value of timber exports from Brazil is approximately \$1 billion per year. Thus, dealing with deforestation is not an easy task. Nearly 40% of wood cut in the Amazon is shipped overseas; the figure was 14% in 1999. Brazil's main markets are: Europe (40%), USA (33%), and China (14%). Although the Brazilian government requires permits for cutting down trees, more than half of the timber exports are illegal.

Professor Granger discussed the sheer loss resulting from deforestation. Based on past data, Professor Granger and co-authors have estimated that more than 50% of the newly-cleared land in the Amazon forest becomes fallow within 5 years, a huge waste of the Amazon forest.

Professor Granger also discussed deforestation related to settlers. Part of the reason for the increase in settlers appears to be increased access to roads. In the Brazilian Amazon, nearly 75% of the deforestation has occurred with 50 km of paved roads. He pointed out that Alexander Pfaff and co-authors (2006) observed a direct relationship between new roads and increased deforestation.

Professor Granger also discussed how to build an econometric model linking important variables measuring the forest with socio-economic, physical, and policy variables using a panel data set of mixed quality. The problem of how to effectively introduce policy variables was given particular consideration. Results from an initial study were presented, but he emphasized that recent considerations suggest that a different model is required.

Professor Granger's presentation was well received by the audience and they also appreciated the fact that a person of his calibre has added his support to this important issue.

The Amazon study discussed in this presentation has been published; the reference is:

Lykke Anderson and Clive W. J. Granger, Estaquio Reis, Diana Weinhold and Sven Wunder (2002). Dynamics of deforestation and economic growth in Brazilian Amazon. Cambridge University Press.

Mervyn J. Silvapulle

4) Belz Lecture Report



Rob Hyndman giving the Belz lecture

The 2006 Belz Lecture was held on Tuesday 24th October 2006 in Old Geology Theatre 1, University of Melbourne. Over 80 people enjoyed this 38th Belz lecture which was presented by Professor Rob Hyndman from the Department of Econometrics & Business Statistics Monash University. The title of his talk was: **"Forecasting and the importance of being uncertain"**.

Professor Hyndman began the talk with a recent news item where a Russian woman is suing weather forecasters for wrecking her holiday. Alyona Gabitova had been promised 28 degrees and sunshine when she planned a camping trip. But it did nothing but pour with rain the whole time, leaving her with a cold. Gabitova has asked the court to order the weather service to pay the cost of her travel.

He then gave a number of amusing forecasting quotes. These included:

"I think there is a world market for maybe five computers." (Chairman of IBM, 1943) and *"We are ready for any unforeseen event which may or may not occur."* (Dan Quayle)

Rob discussed a brief history of forecasting, and gave examples of ancient forecasting methods, such as using a sheep's liver, hallucinogens and vapour-ridden caves in the mountains of Greece. There was even a time when the British Vagrancy Act made it an offence to defraud by charging money for predictions. The punishment was a fine or three months' imprisonment with hard labour.

He claimed that some standard forecasting business practices today are no better than a sheep's liver or hallucinogens. This was demonstrated in the following dialogue:

Budget Maestro utilizes database technology for real-time data collection and reporting. A common interface for all users fosters collaboration and increases the accuracy of data entry. There are no formulas or macros to create, no tedious re-keying of data and no mystery links to chase down and fix. Budget Maestro's built-in "financial intelligence and business rules" builds the formulas for you ensuring 100% accuracy.

Professor Hyndman explained that these methods are largely guesswork, highly subjective, not replicable or testable, and with no possible way of quantifying probabilistic uncertainty. The lack of uncertainty statements leads to false sense of accuracy.

He argued for the importance of statistical modelling in forecasting, and demonstrated the dangers that occur when uncertainty is ignored. In particular, he considered the importance of prediction intervals in providing forecasts, and the dangers of "what-if" scenario projections that are popular in business and government. He outlined some problems involved with data mining prediction and gave the advantages of using stochastic models.

This was demonstrated via two case studies, one on forecasting the Pharmaceutical Benefits Scheme (PBS), the other on predicting the Australian population. An ABC news item had shown that the Federal Opposition had called for tighter controls on drug prices after the PBS budget blew out by almost \$800 million. Rob had found that the Department of Health and Aging were doing all forecasts with the FORECAST function in MS-Excel applied to three points of 10 year old data! Rob tackled this problem with time series models, using automated exponential smoothing state space models applied to about 100 product groups. The methodological tools were developed in 2002 and published in the *International Journal of Forecasting*. The results show that the forecast error is now only a few \$million per year. In the population study, he pointed out distinctions between projections and forecasts, and how better use could be made of prediction intervals with specified probability coverage for population size and relevant derived variables.

He claimed that our traditional emphasis on p-values has been distracting and unhelpful. Rather he sees that most statistical problems are better expressed as prediction problems. Some implications he sees for statisticians include the need to learn what data miners are doing and contribute to the development of their methods, to know (and teach) a lot more computer science, and to teach data mining methods alongside statistical methods in universities. Furthermore he sees that statisticians will place less emphasis on p-values and more emphasis on the predictive ability of models and thinks it will become common to compute prediction intervals for data mining methods. After the talk there was a vigorous question and answer session.

Overall Rob's talk was most entertaining, informative and brilliantly presented. The slides of this presentation are available from <http://www.robhyndman.info/>

Preceding the Belz lecture, Ray Watson was presented with an SSAI Service award from Neville Bartlett. This was presented for his long involvement in SSAI activities, in particular his role as the Technical Editor for the Australian and New Zealand Journal of Statistics from 1998–2005.



Ray Watson (left) and Neville Bartlett

Following the lecture 36 attendees enjoyed a very nice meal and further animated discussions at The University Café in Lygon St.

Tristan Barnett and Brian Phillips

5) Website

Members are reminded that all activities of the SSAI (Vic Branch) are found at <http://www.statsoc.org.au/Branches/VIC/>

Brian Phillips
President SSAI (Vic Branch)
bphillips@swin.edu.au