TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA
UNIVERSITY OF WELLINGTON

New Zealand Statistical Association Conference

2 – 3 September 2009

The
New Zealand
Statistical Association

Welcome to the NZSA 2009 Conference

Welcome to Victoria University of Wellington and to the 60th annual New Zealand Statistical Association Conference, running from 2 to 3 September 2009. Please help us to celebrate this very happy birthday for the NZSA over the next two days!

I am delighted that Professors Matt Wand and Vijay Nair are able to join us this year to deliver their Invited Plenary Talks. Professor Wand has already given a one-day short course on *Semiparametric Regression* here at Victoria on 1 September, which over 60 people attended. Professor Nair will be remaining in New Zealand for another 10 days, as the 2009 Shayle Searle Visiting Fellow in Statistics at Victoria. During that time Professor Nair will give two public lectures, which you are all most welcome to attend; please ask one of the Organising Committee for further details.

I hope you agree that the range of papers and posters on offer is impressive for a two-day meeting, and I trust you will all have a stimulating and enjoyable conference, plus some fun in Wellington.

Thanks are due to the rest of the Organising Committee and to our generous sponsors. Organisers: Richard Arnold, Stefanka Chukova and Ivy Liu (all VUW), Catherine Cameron, Harold Henderson, Roger Littlejohn (all AgResearch), Mike Camden (Statistics New Zealand) and Ian Westbrooke (DOC). Sponsors: Offlode, SAS, VUW School of Mathematics, Statistics and Operations Research, and Statistics New Zealand. Thanks too for administrative support from local staff and students, with special mention for Rowan McCaffery, Christo Muller and Ginny Whatarau (all staff).

If you have any queries concerning facilities at Victoria, please ask one of the local academic staff or students who are at the Conference, identified from the affiliation on their name badges.

Best wishes,

John Haywood
NZSA 2009 Conference Chair
September 2009

CONTENTS

Presenters below are listed in alphabetical order (using last names) *within each type of event*.

Conference Programme

The conference programme for Wednesday 2 and Thursday 3 September is given below.

Contributed talks have been timetabled in alphabetical order within sessions, except in a few specific cases. **The last speaker in each contributed session will be the Session Chair, except in the Student sessions where one of the judges will do that job. Could session chairs please make sure all contributed talks keep to time (up to 18 minutes talk, plus at least 2 minutes for questions – 20 minutes in total)**, since we have a busy schedule and some people may wish to swap between rooms at changeover times during the sessions.

Presenters are responsible for ensuring their talk is 'ready to go' when their session starts: please set up at the latest by the end of the refreshment break before each session. Pdf and powerpoint presentations should both work with no problems, directly from a USB. Connections for personal laptops are available but **if you want to use a Macintosh, please use your own adapter**. Projectors for overhead transparencies will be available in the rooms, if required.

Wednesday 2 September

Time	Event and Location
8.30am onwards	**Registration: Maclaurin building, Foyer**
9.00-9.10am	**Welcome and Housekeeping: CO LT122** Welcome Speech: Pat Walsh, Vice-Chancellor, VUW Housekeeping: John Haywood
9.10-10.10am	**Plenary Talk 1: CO LT122. Chair: John Haywood** Matt Wand: *Variational approximations in semiparametric regression*
10.10-10.30am	**Morning Tea/Coffee: Maclaurin building, Foyer**

	Session A1: CO 118 Time Series Chair: Kit Withers	Session A2: CO LT122 Biostatistics 1 Chair: Esther Meenken	Session A3: CO 119 Student Talks 1 Chair: Jim Young
10.30- 10.50am	Laimonis Kavalieris: *Counting breakpoints in time series*	Rod Ball: *A Virtual Institute of Statistical Genetics*	Cherif Aidara: *The use of auxiliary information in dual frame estimation*
10.50- 11.10am	Claudia Kirch: *Bootstrapping sequential tests for change-point analysis*	James Barton: *New Zealand's GHG inventory – A beginners guide for NZ statisticians*	Brigid Betz-Stablein: *Modelling glaucoma with censored and error prone predictors*
11.10- 11.30am	Norio Watanabe: *Estimation of average trajectory of nearly periodic motion*	Austina Clark: *Estimating species richness and comparing the species similarity using various models*	Keng-Hao Chang: *Robust regression using nonparametric scale normal mixtures*
11.30- 11.50am	Guan Yu Zheng: *Viability of reducing survey frequency using administrative data*	James Degnan: *Bimodal distributions of coalescent times given a gene tree in a species tree*	Chew-Seng Chee: *Semiparametric mixture models for symmetric density estimation*
11.50am- 12.10pm	Kit Withers: *The distribution of the maximum of a first order moving average: The continuous case*	Esther Meenken: *From town to farm: Compost in Canterbury agricultural systems*	Tilman Davies: *On adaptive kernel estimates of spatially dependent disease risk*
12.10- 1.10pm	**Lunch: Maclaurin building, Foyer**		

	Session B1: CO 118 Probability, Regression and Calibration Chair: Walter Davis	Session B2: CO LT122 Statistical Thinking and Applications Chair: Stefanka Chukova	Session B3: CO 119 Student Talks 2 Chair: James Reilly
1.10-1.30pm	Clint Jowett: *Quantifying the division between professional and amateur rugby in New Zealand*	Jennifer Brown: *Reducing uncertainty about uncertainty*	James Dawber: *An applied analysis of bird counts using generalised additive models*
1.30-1.50pm	Frances Krsinich: *Using hedonic regression to assess the housing rentals component of the New Zealand Consumers Price Index*	Rolf Turner: *A dearth of evidence*	Sarojinie Fernando: *Estimation of geographical relative risk by local polynomial regression*
1.50-2.10pm	Dong Wang: *Properties of a random matrix with multivariate normal columns*	Irene David: *Engaging the University 100 level audience – How to teach statistical thinking to students from a mixed background*	Diane Hindmarsh: *The NSW population health survey and small area estimation and analysis*
2.10-2.30pm	Selvanayagam Ganesalingam: *On triangular approximation to symmetric probability distributions*	Kuldeep Kumar: *Bankruptcy prediction using decision tree technique*	Miriam Hodge: *Calculation of fix-rate bias for automated telemetry systems*
2.30-2.50pm	Walter Davis: *Playing with matches: Probabilistic linkage, uncertainty bias and variance*	Stefanka Chukova and Dimitar Christozov: *Analysing truncated warranty data: A stratification approach*	Ting Wang: *Hidden Markov models and mutual information investigation of possible link between GPS measurements and earthquakes*
2.50-3.20pm	**Afternoon Tea/Coffee: Maclaurin building, Foyer**		

	Session C1: CO 118 Longitudinal Data Chair: Ivy Liu	Session C2: CO LT122 Testing and Estimation Chair: Peter Thomson	Session C3: CO 119 Student Talks 3 Chair: Judith Archibald
3.20-3.40pm	Deborah Brunning: *Experiences with the design and analysis of longitudinal data at Statistics New Zealand*	John Haywood: *Here today, gone tomorrow? Exponential reign lengths revealed and explained*	Marissa Isidro: *Intercensal updating of small area estimates*
3.40-4.00pm	John Bryant: *Attrition in the Longitudinal Immigration Survey New Zealand*	Steve Quinn: *A comparison of logistic regression goodness-of-fit statistics when applied to the log link*	Hinrich Kozik: *A case for Q? An application of Q-methodology*
4.00-4.20pm	Reuben Kendall and Mike Camden: *Confidentiality of longitudinal collections*	Robin Willink: *A general construction of a shrinkage confidence interval*	Anna MacDonald: *Mixture of parametric and non-parametric models for extreme values with threshold estimation*
4.20-4.40pm	Chikako van Koten: *Longitudinal analysis of count data: A GEE approach*	Lingyun Zhang: *A note on Bartlett's M test for homogeneity*	Aghababaei Jazi Mansour: *Some extensions of the geometric distribution*
4.40-5.00pm	Ivy Liu: *A hybrid method for longitudinal data*	Peter Thomson: *Mixed methods for fitting the GEV distribution*	Katharina Parry: *Traffic network-based modelling*
5.05pm	**NZSA Annual General Meeting: CO LT122**		
7.00pm	**Pre-dinner drinks: The Skyline restaurant, 1 Upland Rd, Kelburn**		
7.30pm	**Conference Dinner: The Skyline restaurant, 1 Upland Rd, Kelburn**		

Thursday 3 September

Time	Event and Location		
8.15-9.00am	**Young Statisticians' Breakfast/Networking: Maclaurin building, Foyer**		
	Session D1: CO 118 Bayesian Statistics Chair: Richard Arnold	**Session D2: CO LT122 NZ Society and Economy Chair: Steven Johnston**	**Session D3: CO 119 Biostatistics 2 and Analytics Chair: Alastair Scott**
9.00-9.20am	Rahul Mukerjee: *Highest posterior density regions based on empirical-type likelihoods: Role of data-dependent priors*	Mike Doherty: *Sampling errors on the six degrees of separation*	Yuichi Hirose: *Efficiency of profile/partial likelihood in the Cox model*
9.20-9.40am	Magnus McGee: *Bayesian calibration in Microsimulation of Disease and Cancer in New Zealand (MoDCoNZ)*	Brian Easton and Ryan You: *Demographic determinants of life satisfaction*	Evan Stubbs: *Feedback from the field: How our graduates are letting down the profession*
9.40-10.00am	Richard Penny: *Imputation variance estimation for the Accommodation Occupancy Survey*	Brian Easton, Ryan You and Sally Casswell: *Life satisfaction and drinking: A preliminary investigation*	Saibal Chattopadhyay: *Exponential clinical trials: Sequential comparison under asymmetric penalty*
10.00-10.20am	Finlay Thompson: *Building a two stage Bayesian dolphin capture model with JAGS*	Jamas Enright: *The effects of health and wealth on the labour supply and retirement decisions of older New Zealanders*	Alastair Scott: *Non-response in case-control studies*
10.20-10.40am	Richard Arnold: *Capture-Recapture estimation using finite mixtures of arbitrary dimension*	Steven Johnston: *Using statistical methods to answer policy questions: Sole parents, employment incentives and benefit receipt*	
10.40-11.00am	**Morning Tea/Coffee: Maclaurin building, Foyer**		
11.00-12noon	**Plenary Talk 2: CO LT122. Chair: John Haywood** Vijay Nair: *Process monitoring, anomaly detection, and beyond: New directions and applications*		
12.00-1.00pm	**Lunch: Maclaurin building, Foyer**		

1.00-2.00pm	**Statistics Education parallel workshops, first stream: CO 118**	**Plenary session on the future of the Australian and New Zealand Journal of Statistics: CO LT122 Chair: Ian Westbrooke**	**Statistics Education parallel workshops, second stream: CO 119**
1.00-1.30pm	Maxine Pfannkuch: *Comparative statistical reasoning: Some essential dialogues*		Tim Burgess: *The challenges of using statistical investigations in primary school classrooms*
1.30-2.00pm	Paul Bucknall, Mike Camden and Nathaniel Pihama: *A website of statistical riches*		Derek Smith: *Probability trees and two way tables*
	Statistics Education talks: CO LT122 Chair: Mike Camden		
2.00-2.20pm	Mike Camden: *New resources for learning of the Curriculum's statistics*		
2.20-2.40pm	John Harraway (presented by Mike Camden): *A new teaching resource for Year 13*		
2.40-3.00pm	Kevin Burns: *How smart are birdbrains? Number sense in the New Zealand robin*		
3.00-3.20pm	**Afternoon Tea/Coffee: Maclaurin building, Foyer**		
	Statistics Education talks: CO LT122 Chair: Mike Camden		
3.20-3.40pm	Robin Averill: *Learning in Statistics: Perspectives of secondary school students*		
3.40-4.00pm	Alasdair Noble and Anne Lawrence: *One good tern...*		
	Statistics Education parallel workshops, first stream: CO 118		**Statistics Education parallel workshops, second stream: CO 119**
4.00-5.00pm	Robin Averill: *Ideas for teaching statistics in junior secondary school classrooms*		Sandi Tait-McCutcheon: *Framing Statistics and Probability Level 1 to 5*

Matt Wand

School of Mathematics and Applied Statistics, University of Wollongong

Variational approximations in semiparametric regression

Variational approximations are a body of analytic procedures for handling difficult probability calculus problems. They have been used extensively in Statistical Physics and Computer Science. Variational approximations offer an alternative to Markov chain Monte Carlo methods and have the advantage of being faster and not requiring convergence diagnoses, albeit with some loss in accuracy. Despite the growing literature on variational approximations, they currently have little presence in mainstream Statistics. We describe recent work on the transferral and adaptation of variational approximation methodology to contemporary Statistics settings such as generalised linear mixed models and semiparametric regression. This talk represents joint research with Professor Peter Hall and Dr John T. Ormerod.

Vijay Nair

Department of Statistics

Department of Industrial & Operations Engineering

University of Michigan, Ann Arbor

Process monitoring, anomaly detection, and beyond:

New directions and applications

There is a huge literature on anomaly detection dealing with statistical process monitoring in manufacturing, engineering process control, and change-point detection. More recent work includes methods for multivariate data, time series, integrating statistical and engineering process control and so on. In this talk, we will provide a brief review of these methods and describe recent developments motivated by new applications. These include various issues in advanced manufacturing, fraud detection in telecommunications and finance, fault detection in sensor networks, monitoring quality of service in communications networks, and biosurveillance.

NZSA 2009 Abstracts for Contributed Talks: Students

Abstracts appear alphabetically, ordered by the speaker's last name

Cherif Aidara

Victoria University of Wellington

The use of auxiliary information in dual frame estimation

In this paper we propose extending the dual frame estimators proposed by Hartley (1962) and Mecatti (2007) to incorporate auxiliary information. Techniques known to be efficient in incorporating auxiliary information in the estimation procedure are used. These techniques are namely the generalised regression and calibration. The resulting dual frame estimators and their variances are discussed in this paper.

Brigid Betz-Stablein

Massey University

Modelling glaucoma with censored and error prone predictors

Glaucoma is the second leading cause of blindness in the world, accounting for 12.3% of blindness worldwide. Glaucoma is caused by a build-up of fluid in the eye, which can lead to permanent, irreparable vision loss. While glaucoma subjects' eyes are more likely than normal eyes to lack spontaneous venous pulsation (SVP), this can be induced by applying pressure to the eye using an ophthalmodynamometer. Recently, research undertaken by The Lions Eye Institute (Perth, WA) has suggested that the pressure measured, ophthalmodynamometric force (ODF), is related to the severity of glaucoma. The amount of data available is limited; therefore it is important to employ efficient statistical models to investigate this relationship. Recorded ODF values are subject to measurement error, which is complicated by heteroscedastic variation and the fact that the data is also censored at zero. We employed hierarchical models within a Bayesian framework to investigate the effect of change in errorless ODF on increased excavation of the optic disk, a measure of glaucoma progression. Our results showed that size of the effect of ODF on glaucoma progression increased compared to models that do not account for measurement error and censoring. Therefore change in ODF may provide more information than originally thought to assist in glaucoma prognostication, and may play a larger role in assessing the likelihood of glaucoma progression.

Keng-Hao Chang

University of Auckland, Department of Statistics

Robust regression using nonparametric scale normal mixtures

A standard assumption for linear regression is that the random noise has a normal distribution. However, for some real-world data this assumption may not necessarily hold and the model can be badly fitted, e.g., when the random noise follows a distribution with heavier tails than the normal or when there exist outliers. In this talk, I will describe a method that uses nonparametric scale normal mixtures to solve this problem.

Chew-Seng Chee

Department of Statistics, University of Auckland

Semiparametric mixture models for symmetric density estimation

We present a general semiparametric framework based on mixtures for univariate symmetric density estimation and propose a semiparametric mixture symmetric density estimator (spsym). The performance of the estimator hinges on appropriate choice of the tuning parameter. To this end, we introduce a simple strategy for selecting the tuning parameter in practical implementations. Since the spsym is essentially a semiparametric mixture model, this allows us to take advantage of the CNM-MS algorithm used to fit semiparametric mixture models and with minor modification the algorithm can be used for the spsym computation. A simple real example shows that the mixture-based method provides an attractive complement to the traditional kernel-based method. The performances of the mixture-based and kernel-based methods are illustrated through a simulation study.

Tilman Davies

Massey University

On adaptive kernel estimates of spatially dependent disease risk

Kernel smoothing is a well known technique for the estimation of relative risk based on point locations of disease cases and sampled population controls contained within a defined geographical region. Most applications of this methodology have made use of the fixed-bandwidth version of the kernel estimator in order to obtain estimates of the requisite densities. A more intuitive approach is to utilize a variable smoothing parameter with which we are able to reduce the bandwidth in areas of high point clustering to provide more detail and increase it to 'smooth over' isolated observations in regional areas, a technique shown to yield certain theoretical benefits for the individual densities. Despite these potential benefits, the increased complexity involved with the adaptive estimator has left a number of technical issues such as bandwidth selection and edge-correction that warrant additional attention. Problems also present themselves should we wish to make use of the adaptive risk surface for detection of significant 'hotspots' of disease risk due to excessive computation time. The technical issues to do with the implementation of the adaptive surface are addressed and we develop a computationally inexpensive approach for highlighting anomalous sub-regions of risk on the adaptive surface. The performance of both fixed and adaptive surfaces in terms of integrated square error is then examined based on results from simulation experiments.

James Dawber

University of Canterbury

An applied analysis of bird counts using generalised additive models

This is joint work with Jennifer Brown (University of Canterbury) and Ian Westbrooke (DOC)

Bird counts are regularly collected across New Zealand in an attempt to understand the abundance of certain bird species. Analysing such data accurately becomes an important role in describing the nature and dynamics of these bird populations. The application of Generalised Additive Models (GAMs) to such datasets is one possibly effective way in producing good models. The application of GAMs on bird counts was explored using real life data, as part of a restoration project lead by DOC in Nelson Lakes National Park. The advantages of using such models on count data are explained, along with the issues and drawbacks that were encountered in the modeling process.

Sarojinie Fernando

IFS/Statistics, Massey University, Palmerston North

Estimation of geographical relative risk by local polynomial regression

A common problem in epidemiology is the estimation of spatial variation in risk associated with a disease. In this talk, we focus our attention on the estimation of a relative risk function using data on the locations of cases and controls in geographical epidemiology. This can be obtained by using a ratio of two bivariate kernel density estimates (the density ratio method). A novel alternative is to use local polynomial regression to estimate the relative risk function. We have studied the properties of the latter method and derived the asymptotic bias and variance of local polynomial estimates. The performance of estimators based on the above mentioned methods are compared through a Monte Carlo simulation study and illustrated using five synthetic problems. An application on disease in myrtle trees is presented.

Diane Hindmarsh

CSSM, School of Mathematics and Applied Statistics, University of Wollongong

The NSW population health survey and small area estimation and analysis

This is joint work with David Steel and Ray Chambers, both CSSM

The CATI-based population health survey program run by NSW Health publishes annual estimates at the state level and for the eight health areas for a large number of health related measurements. Each of the eight health areas is geographically and socially diverse, and a single estimate for an entire area ignores this fact. Direct estimators at the Local Government Area level, or for age-sex domains at the health area level, have relatively high standard errors, which affects their use in health evaluation and planning. There is a desire to use small area estimation methodologies to provide estimates that have greater precision, but two questions need answering: which of the many methods is best suited to this specific situation, and, how do we educate the end-users about the resultant estimates so that they use them with confidence rather than revert to the direct estimates. This paper will outline the NSW Health Survey program and how we are answering these two questions.

Miriam Hodge

University of Canterbury

Calculation of fix-rate bias for automated telemetry systems

GPS and other radio tracking equipment are becoming more widely used by researchers for modelling animal habitat. In a typical monitoring program an animal will be fitted with a tracking collar. This tracking collar will fix the animal's location at a set time interval. These fixes of the animal's location can then be cross referenced on a digital map (GIS) containing habitat information and the animal's preferred habitat can be modelled.

Care must be used in modelling the habitat because radio tracking collars have different transmission probabilities in different habitats. The habitat observations are biased towards habitats that allow good transmission. One way to minimise this bias is to weight observations by a measure of transmission quality.

Researchers have attempted to estimate the detection weighting by placing stationary collars in the study area and recording the fix-rate. The results of these studies are unsatisfactory because stationary collars do not account for animal movement and behaviour. Johnson (1998) used a surrogate for stationary collars by analysing 6 hour time periods where the animal was relatively stationary. We will develop this method further by incorporating the non-stationary sites in the detection rate calculation.

Marissa Isidro

IFS, Massey University

Intercensal updating of small area estimates

The main objective of my research has been to develop a feasible and theoretically sound intercensal updating method for small area estimates of poverty statistics in third world countries. My proposed method is called Extended Structure PREserving Estimation (ESPREE) method, this approach is an extension of the classical small area estimation technique called structure preserving estimation (SPREE) method – an approach to small area estimation based on a categorical data analysis framework. ESPREE is structured within a generalized linear model (GLM) framework and uses information from the most recent survey and pseudo-census (replicated sets of the census data) data to generate updated small area estimates.

The World Bank in collaboration with the National Statistical Coordination Board in the Philippines has conducted an intercensal updating project using a methodology based on the widely implemented small area estimation methodology for poverty measures in third world countries called ELL (Elbers, Lanjouw and Lanjouw). The estimates generated from the ELL and the ESPREE updating methods were compared and substantial differences were observed not only at the small area level (municipality) but also at higher levels of aggregation e.g., provincial and regional levels. The ESPREE method seemed to work better than the ELL updating method in that the ESPREE method generated unbiased estimates. An in-country validation exercise was conducted allowing for better assessment of acceptability and consistency of the estimates generated with the available indicators at the municipal level as well as with the expert opinion of key informants.

Hinrich Kozik

Victoria University of Wellington

A case for Q? An application of Q-methodology

Q-methodology (or simply 'Q') provides a foundation for the systematic study of subjectivity, a person's viewpoint, opinion, beliefs, attitude, and the like. In Q-methodology people (and not their traits) are subject to correlation and factorisation: instead of scoring a large number of people on a small number of variables, Q-methodology requires a small number of people to rank-order a relatively large number of test items, such as statements, and then factorises these scores. Although there is plenty of evidence of controversy and peer criticism regarding Q-methodology (particularly up until the late 1960s) it is now being widely adopted as a means of investigation for attitude measurement, particularly in North America. Using an application of Q-methodology (my exploratory MA research on impacts of telecommuting on people with disability) the talk will introduce the audience to the basics of Q.

Anna MacDonald

University of Canterbury

Mixture of parametric and non-parametric models for extreme values with threshold estimation

The traditional asymptotically motivated extreme value model for the tail(s) of distributions (exceedances over high thresholds) is the Generalised Pareto distribution. Substantial uncertainty can be introduced to tail estimates due to the selection of the threshold. Typically, somewhat subjective threshold choices are made using graphical tools. Recently, various mixture type models have been proposed for the entire distribution function, simultaneously capturing the bulk of the distribution with the flexibility of the Generalised Pareto for the upper/lower tails. These mixture models either explicitly include the threshold as a parameter to be estimated, or somewhat bypass this choice by the use of smooth transition functions between bulk and tail models.

We introduce a new mixture model, based on using smoothing techniques to capture the bulk of the distribution with the inclusion of the threshold as a parameter to be estimated, using Bayesian techniques. Comparisons of the model are made using both traditional techniques and other mixture models.

The primary goal is exploring mixture models and their application for modelling high frequency physiological measurements for preterm babies. Via the estimation of suitable quantiles, we are looking to refine our understanding of "normal ranges" for the level and variability of these measurements. The proposed model is applied to blood oxygenation from preterm babies in the neonatal intensive care units at Christchurch Women's Hospital, New Zealand.

Aghababaei Jazi Mansour

IFS/Statistics, Massey University, Palmerston North

Some extensions of the geometric distribution

We shall consider some extensions of the geometric distribution which are usually used in life time distributions. The discrete Weibull distribution as a discrete analogue of the standard continuous Weibull distribution was proposed in the literature to meet the need of fitting discrete time reliability and survival data sets. Its properties were studied and methods of estimation of its parameters were also investigated by various authors. Analogous to its continuous counterpart, the discrete Weibull does not provide a good fit to data sets that exhibit non-monotonic hazard rate shapes. So, we propose a discrete inverse Weibull distribution which is a discrete version of the continuous inverse Weibull variable. Then, we shall discuss their structural properties such as modality and divisibility. It is shown that the hazard rate function can attain a unimodal or monotonely decreasing shape for certain values of parameters.

Katharina Parry

Massey University

Traffic network-based modelling

Network-based models of traffic systems provide vital insights for road planning and management. The focus here is on MCMC inference for models of networks with a tree structure based on traffic count data observed on a set of network links.

Many of the model parameters are concerned with travellers' route choices, for which inference would be straightforward if we were to observe path flows. It is therefore natural to sample these latent variables within the MCMC algorithm. A critical problem is that there are very large numbers of possible route flows that could have led to an given observed set of link counts. It becomes essential to find a suitable sampling scheme which avoids the evaluation of the full set of feasible route flows.

We present a solution using first-order Markov model of traveller behaviour to generate candidate route flows in a Metropolis-Hastings sampler.

Ting Wang

IFS/Statistics, Massey University

Hidden Markov models and mutual information investigation of possible link between GPS measurements and earthquakes

Geodetic anomalies preceding large earthquakes have long been of interest due to well-documented pre-earthquake deformation rate changes observed before continuous GPS stations were widely deployed in the early 1990s. These GPS measurements provide a good opportunity for scientists further investigating pre-, co- and post-seismic deformation anomalies, but there is much 'noise' that needs to be filtered out of the observations. Different variables of GPS deformation measurements are examined and fitted by a hidden Markov model using a multivariate normal observation density. The model classifies the data into different categories of states. Each state of data may suggest particular dynamics. The mutual information (a measure of the amount of information that one random variable contains about another) between each state of the most likely state sequence from the HMM analysis, and the earthquake occurrence data is examined, using data from central North Island, New Zealand. We discuss the detected states including the precursory signals and relate them with different dynamics. After this we discuss a possible way of declaring a "Time of Increased Probability" (TIP) for the considered region. Consequently, probability forecasts in time based on a logistic regression model are investigated.

NZSA 2009 Abstracts for Contributed Talks

Abstracts appear alphabetically, ordered by the speaker's last name

Richard Arnold

Victoria University of Wellington

Capture-Recapture estimation using finite mixtures of arbitrary dimension

Reversible Jump MCMC methods are used to fit Bayesian capture-recapture models incorporating heterogeneity in individuals and samples. Heterogeneity in capture probabilities comes from finite mixtures and/or fixed sample effects allowing for interactions. Estimation by Reversible Jump MCMC allows automatic model selection and/or model averaging. Priors on the parameters stabilise the estimates and produce realistic credible intervals for population size for over-parameterised models, in contrast to likelihood based methods. To demonstrate the approach we analyse the standard Snowshoe Hare and Cottontail Rabbit data sets from ecology, and a reliability testing data set.

Rod Ball

Scion

A Virtual Institute of Statistical Genetics

The Virtual Institute of Statistical Genetics (VISG) is a FRST funded research program involving Universities and Crown Research Institutes, that has just completed its first year of operation. I will describe progress on current projects (Large Datasets and Polyploids) and plans for the next project (Experimental Designs).

Whole genome prediction of genetic values is already being applied to livestock in New Zealand. The VISG large datasets project is developing methods for whole genome association mapping and prediction of genetic values. Currently we are working on a Bayesian Markov chain Monte Carlo (MCMC) method for fitting associations using a multi-category prior (with 3 categories, with separate variance parameters for each category) and block updates for SNP effects. The modelling approach allows for low prior probabilities for non-negligible SNP effects (necessary for a $p \ll n$ problem), and a non-normal mixture distribution for effect sizes. Special attention is given to algorithms for improving and diagnosing MCMC convergence to avoid problems with existing QTL or whole genome MCMC methods. Also important are algorithms to effectively handle large datasets with currently of the order of 1,000,000 SNP markers genotyped per individual.

A number of important horticultural, crop, and forage species are polyploids. Existing QTL mapping in polyploids is limited to specific marker types and segregation patterns, and inference is limited. The VISG polyploids project is developing methods for QTL mapping in polyploids which make full use of available marker information and enable multi-locus Bayesian inference of the genetic architecture.

Polyploids have 2 or more sub-genomes resulting in (e.g. for an allo-tetraploid) 4 or more alleles at each locus each of which could have been inherited from one of 8 grand-parental chromosomes. Markers are rarely fully

informative, so that the statistical method needs to contend with considerable and variable amounts of missing information. This is being done by integrating peeling and conditional peeling with a Bayesian QTL mapping method.

Experimental design has been a neglected area in genomics, with even large scale international projects lacking power to detect any but the largest effects with posterior odds greater than 1. The VISG experimental designs project will develop experimental designs with sufficient power to detect genomic associations, with sufficiently high Bayes factor to overcome the low prior odds for genomic associations, and utilising design and analysis options available in various species (e.g. clonal replication and spatial analysis).

James Barton

James P Barton & Associates Limited

New Zealand's GHG inventory – A beginners guide for NZ statisticians

The compilation of annual national greenhouse gas (GHG) inventories in the Environment sphere is a complex estimation exercise which has become as significant as National Accounting is in the Economy sphere. For New Zealand it is now a requirement under the Kyoto Protocol (KP) to the United Nations Framework Convention on Climate Change (UNFCC). The presentation will briefly introduce key concepts used in GHG inventory reporting, how the UNFCC GHG Inventory is currently completed for New Zealand and the implications for New Zealand if it is done poorly. Trends and uncertainties will be illustrated for key sectors and the implications of the Compliance Equation will be outlined. In this way it is hoped that some of the statistical alchemistry used in GHG accounting and reporting will become a little clearer for NZ statisticians.

Jennifer Brown

University of Canterbury

Reducing uncertainty about uncertainty

This is joint work with Timothy J. Robinson, University of Wyoming

What does uncertainty mean? As statisticians we use this term regularly, but to a biologist it can be rather unsettling to hear there is "uncertainty" in the results from their study.

One way to explain uncertainty, so there is no uncertainty about what it is, is to illustrate (by way of simulation) where it has come from. We discuss three case studies where we have used a simulation approach to illustrate the pathways and accumulation of uncertainty. We discuss uncertainty in penguin counts, in river bird surveys and in forest biodiversity sampling.

The end result of illustrating the final level of uncertainty to field-biologists has always been the same, with the biologist saying "I wish I hadn't asked"! Once they have dealt with that reaction, the exercises have been useful to help scope where future survey designs can be modified to reduce uncertainty.

Deborah Brunning

Statistics New Zealand

Experiences with the design and analysis of longitudinal data at Statistics New Zealand

Prior to the turn of the century, Statistics New Zealand had little experience with the design and analysis of longitudinal data. However over the last decade, the interest of policy makers in New Zealand, like their counterparts in many other countries, has identified a need for more information to enable them to study patterns and dynamics beyond what is achievable with repeated cross-sectional snapshots. To respond to this need over the past 10 years Statistics New Zealand has: designed and run 7 waves of an 8 wave survey to measure income and employment dynamics (known as SoFIE); conducted, in partnership with the Department of Labour, 2 waves of a 3 wave longitudinal survey to measure migrant's settlement experiences in New Zealand (LISNZ); developed a longitudinal business database, by combining data from a number of sources; and developed a dataset which enables us to examine the longitudinal patterns and dynamics of both employers and employees using administrative data from the taxation system (LEED).

In this paper we will discuss our experiences with these developments. There have been both achievements and significant challenges encountered in this work, in areas such as design and implementation of collection methodologies, use of computer assisted methods and confidentiality and data access. We will discuss how these experiences are influencing our thinking about future longitudinal data collections in Statistics New Zealand.

John Bryant

Statistics New Zealand

Attrition in the Longitudinal Immigration Survey New Zealand

This is joint work with Frances Krsinich, Statistics New Zealand

The Longitudinal Immigration Survey: New Zealand (LisNZ) is a uniquely rich source of data on migrants' labour market and settlement outcomes in New Zealand. But, like all longitudinal surveys, it is subject to attrition. Approximately 14 percent of respondents who were interviewed in wave 1 could not be re-interviewed in wave 2. We investigate whether this attrition leads to selection bias in simple cross-sectional models using LisNZ data. We apply two closely-related tests: (i) we examine whether attrition in wave 2 is related to outcomes in wave 1, after controlling for standard explanatory variables; and (ii) we examine whether the relationship between outcome and explanatory variables differs between non-attritors and the complete sample. Both tests suggest the existence of selection bias. These biases are nevertheless small compared with the size of the coefficients, especially when wages are used as the outcome variable. The small size of the biases means that, at current attrition rates, the LisNZ sample essentially remains representative of its target population.

Saibal Chattopadhyay

Indian Institute of Management Calcutta

Exponential clinical trials: Sequential comparison under asymmetric penalty

Suppose that two divisions of the same pharmaceutical company have been independently studying response times for two comparable treatments via separate clinical trials under identical protocols. Each division has been experimenting with just one treatment. The response times in each case follows a two-parameter exponential model, with all parameters unknown. The physical separation of these two divisions is an important input, and sequential design and allocation of units to these two treatments is ruled out. We consider the case where the two divisions independently estimate the minimum guarantee times (i.e., the thresholds) under some asymmetric linear-exponential (linex) loss function having unequal penalties for over and under estimation. Individual divisions have certain notion of 'fixed-precision' (such as bounded risk point estimation) tied with their individual problems. In the absence of fixed-sample size procedures, each division independently performed some sequential experimentation, and came up with independent stopping rules. The present study attempts to combine the results obtained from these two independent sequential studies and proposes to estimate the difference of these two thresholds under a similar asymmetric loss function. Asymptotic expansion of the sequential risk is derived under a general class of stopping rules. Examples of combining specific sequential methodologies such as purely sequential, accelerated sequential etc. are provided.

Stefanka Chukova and Dimitar Christozov

Victoria University of Wellington & American University in Bulgaria

Analysing truncated warranty data: A stratification approach

Warranty data is of considerable interest to corporations for several reasons. Warranty claims are a liability incurred at the time of sale and represent a cost of doing business, so forecasting those costs is of interest. For engineers a secondary, but important, use of warranty data is to assess the reliability of products in the field. A third characteristic of warranty is that it is a product attribute valued by customers and affecting their buying decisions. For example increasing warranty coverage may attract more buyers but also increase servicing costs.

Vehicle age is known at all times because sales records are retained. It is also becoming technically feasible to track mileage accumulation on all vehicles in the field, but this is currently not a common practice for cost and privacy reasons. For vehicles that generate a warranty claim, the mileage at the time of a warranty repair are recorded at the dealership and included in the warranty database. Thus from a modeling standpoint we have two usage measures (age and mileage), but one of them (mileage) is incompletely observed. As is commonly done we model warranty claims as recurrent events from a repairable system. Also, we take a nonparametric approach because sample sizes are large. We note however that warranty forecasting, which requires extrapolation beyond the oldest age/mileage in the field, requires either a parametric model or the incorporation of past-model data on older vehicles. We deal explicitly with the problem of incomplete mileage information and also with the problem that repairs made beyond the age or mileage limits will not be part of the warranty database.

Usually a simple linear mileage accumulation model is adopted to estimate the number of units at risk at any given time from the incomplete mileage data. For mileage accumulation, we relax the linearity assumption, based on the information from the last warranty claim, by proposing a piece-wise linear model with nodes occurring at the observed mileages corresponding to the warranty repairs. For the piece-wise model, we use all claims in the database to characterise the mileage accumulation.

The main focus of this study is to provide an overview of different stratification approaches in analysing warranty data. These approaches are characterised by the choice of an appropriate integral measure, which depends on the usage measures (mileage and age) of the vehicles, and imposes a particular discretisation on the warranty data. The goal of these stratifications is to extract information needed for the evaluation of several warranty characteristics, such as the mean cumulative cost function, the rate of depletion of the warranty resource as well as to detect useful trends in these characteristics.

Austina Clark

University of Otago

Estimating species richness and comparing the species similarity using various models

It has long been realized by ecologists that it is almost impossible to get a complete census of all species within a community with limited resources. It is also difficult to compare the shared species using the finite resources, particularly with communities with large species richness and a large portion of rare species. (Colwell & Coddington 1994; Colwell et al 2004; Magurran 2004).

A total of 1200 samples (via quadrats) were collected over 4 years from 1990 to 1993 in the South Island, New Zealand. There were 10 sites involved here. The ecologist divided each site into 3 regions and then allocated three treatments randomly to these regions. These were treatment 1 (T1) where there were both rabbits and sheep on the region, treatment 2 (T2) where there were rabbits but sheep were removed, treatment 3 (T3) where both rabbits and sheep were removed. Ten quadrats were placed in each region and the ecologist collected samples from the same quadrat each year from 1990 to 1993.

The 10 sites are regrouped into three areas according to the plants growing history, A1 is the land which is covered with tall tussocks, A2 is for Hieracium and A3 is for Bare land.

Here we will apply a few methods and computer software (SPADE) is used to find the species richness in a single community. Next we use the multiple-community diversity measure to compare the similarity (Chao et al, 2008) among the areas as well as the three treatments.

Irene David and Jennifer Brown

University of Canterbury

Engaging the University 100 level audience:

How to teach statistical thinking to students from a mixed background

At the University of Canterbury we have undertaken a major change in the way we deliver the 100 level statistics service course. The new-look course emphasises statistical thinking over methods and computation. The catalyst for change was a recognition that most (~99%) of the students take the course for general interest and the support of other majors, and very few are planning to major in statistics. In the process of managing the change we asked: "Given a limited time frame for delivery, what are the essential statistics skills your bank manager/car salesmen/primary school teacher needs to have?" Ultimately we moved from teaching the statistics skills needed for a statistician to teaching those skills needed for a manager/scientist/'informed consumer' in today's society.

The way the course is now delivered provides lesser emphasis on lectures as the main way of delivering the course (and students' learning experience) and greater on mastery: computer based tutorials (using a Learning Management System), use of Excel's functionality as the ubiquitous tool and on-line skills testing; with written assignments to ensure synthesis and statistical thinking.

With more than 1100 students enrolled on the three occurrences of the course in 2009 we have some interesting observations and experiences to share from the experience to date. Students have been given the opportunity to provide input to surveys at three points in the course: at the outset, to provide a large data set to use in teaching and learning, to engage the students; for formative feedback half way through and for summative feedback at the end of the course. The students' ownership of their learning, opportunities to give feedback, work on-line at their own pace and in their own time, take tutor help if needed and ask for further learning resources moved the focus towards student-centred learning and away from passive to active participation.

Walter Davis

Statistics New Zealand

Playing with matches: Probabilistic linkage, uncertainty bias and variance

A great deal of work (Chambers 2008; Lahiri-Larsen 2005; Scheuren-Winkler 1993) has been done on developing unbiased estimators for probabilistically linked data. This talk will focus on simulations performed to identify thresholds for acceptable matches.

James Degnan

University of Canterbury

Bimodal distributions of coalescent times given a gene tree in a species tree

In evolutionary biology, trees of species relationships are often used to depict which species are most closely related, and such species trees can be used as a basis for classifying species. Genes sampled from different species also have tree-like evolutionary histories (gene trees) that often have similar branching patterns to that of the species tree. Relationships between gene trees and species can be modeled probabilistically to determine quantities such as the expected time to coalescence (i.e., common ancestry) for two gene lineages given the species-level divergence times and population sizes for the species being sampled. In this talk, I show how the distribution of a coalescent time can be obtained by integrating over a joint distribution of gene trees and coalescent times for cases with more than two lineages. This marginal distribution for the coalescent time can be regarded as a mixture distribution and can be multimodal when population sizes are not constant.

Mike Doherty

Statistics New Zealand

Sampling errors on the six degrees of separation

I will describe a problem on sampling on a small-world network. The network concerned is constructed from Statistics New Zealand's Linked Employer-Employee data set. Because of the size of the network, the average degrees of separation was worked out for a sample of nodes, taken in a certain way. A variance estimate turns out to follow by standard formulae and careful book-keeping.

Brian Easton and Ryan You

SHORE, Massey University

Demographic determinants of life satisfaction

One of the influences on a person's satisfaction with life (or happiness) is each person's personal characteristics. These characteristics interact (so older women are likely to be widows; each characteristic has a different effect in contrast to youth, maleness, and married). This paper reports on the results of a 3000+ respondent survey, in which the effect of each characteristic is measured, controlled for all other characteristics.

Brian Easton, Ryan You and Sally Casswell

SHORE, Massey University

Life satisfaction and drinking: A preliminary investigation

A survey, with 3000+ respondents, asked about life satisfaction and alcohol drinking activities. So does alcohol contribute to life satisfaction ... or not?

Jamas Enright

New Zealand Treasury

The effects of health and wealth on the labour supply and retirement decisions of older New Zealanders

Health status is an important element in the decision to continue working or retire among older workers. Given the demographic projections for the next four decades, there will be increasing opportunities for older workers to remain in the workforce. However, an individual's decision is likely to be influenced by both their health status and their accumulated wealth. This study analyses the influence of health and wealth on the decision to participate in the labour force amongst older New Zealanders, aged 55 to 70. It is based on the first wave of data collected in a longitudinal survey of Health, Work and Retirement conducted by researchers at Massey University.

The study employs a range of measures of health including the results from the international Short Form (SF36), self-reported health status and the prevalence of chronic illness. Regardless of the measures tested, a significant reduction in labour force participation is associated with poorer health status. It is widely recognised that health status itself may be partly determined by labour market characteristics. Attempts to deal with this statistically were not successful. Perhaps surprisingly, wealth did not appear to be strongly related to the decision to retire. A marked fall in participation is associated with the receipt of New Zealand Superannuation, arguably masking the effect of privately held forms of retirement wealth.

Selvanayagam Ganesalingam

IFS, Massey University

On triangular approximation to symmetric probability distributions

The triangular distribution is shown to be a reasonable approximation for the symmetric distributions such as the normal and the t-distribution. In the case of the t-distribution, the approximation improves as the degree of freedom increases. This paper provides a theoretical basis for the use of the triangular approximation, and simulated examples to demonstrate the practical application of this approximation.

John Haywood

Victoria University of Wellington

Here today, gone tomorrow? Exponential reign lengths revealed and explained

This is joint work with Estate Khmaladze and Ray Brownrigg, both VUW

Human lifetimes have an increasing failure rate, or force of mortality; as people age they are more likely to die. We consider the succession of Chinese Emperors from 221 BCE to 1911 CE and show that, unlike lifetimes, their reigns ceased at a constant rate, unaffected by elapsed time or by accumulated social, political or economic tensions. Reign lengths of the 'Sons of Heaven' followed the same memoryless exponential distribution for over two millennia; the 'half-life' of an emperors' reign was 10 years. Thus becoming emperor influenced the duration of remaining life and, surprisingly, length of rule was independent of age at ascent.

To test for exponentiality of reign lengths we use a distribution-free goodness-of-fit test applied to a transformed empirical process, with non-trivial power against alternatives of any type. We also use two tests specifically targeted at an alternative explanation of a mixture of remaining life distributions. Building on certain asymptotic results in the theory of stochastic processes, we propose an explanation of the observed exponential reign lengths, with fundamental implications for the nature of rule: even stable and civilised societies, like Imperial China, do not require nor imply stability of their ruler.

Yuichi Hirose

Victoria University of Wellington

Efficiency of profile/partial likelihood in the Cox model

This presentation shows equivalence and efficiency of the partial likelihood and profile likelihood estimator in the Cox regression model using the direct expansion of the profile likelihood. This approach gives an alternative proof to the one illustrated in Murphy and van der Vaart (2000).

Steven Johnston

Ministry of Social Development

Using statistical methods to answer policy questions:

Sole parents, employment incentives and benefit receipt

The Working for Families policy reforms introduced substantial changes to in-work incentives and financial support for families with dependent children. In this talk we describe one of a number of pieces of work to evaluate the impact of these reforms.

We used (discrete-time) survival analysis to examine changes in the length of time that sole parents spent on benefit and changes in the length of their subsequent spells off benefit. We describe some of the challenges we encountered, including that some individuals had more than one spell of benefit receipt, and finding suitable ways to present the results for a policy audience. Our analysis used a new research dataset of linked MSD and Inland Revenue administrative data that allows us to track families on a monthly basis from 2003 onwards.

Clint Jowett

Offlode / Massey University

Quantifying the division between professional and amateur rugby in New Zealand

This is joint work with Paul J. Bracewell, Daniel C. Walsh and Don G. R. Forbes

Team Lodeings are a useful rating method for measuring intra-competition team performance. This method is based on relative points differential. However, these ratings are not suitable for use when inter-competition matches are played and the ability levels on the distinct competitions cannot be assumed to be equivalent. We develop a method for calibrating intra-competition ratings so that they become meaningful inter-competition ratings. We test and validate our approach using the American domestic gridiron competition (NFL) before applying the results to New Zealand domestic first class rugby. Using the calibrated inter-competition Team Lodeings we will outline how the game of rugby has evolved in New Zealand over the last 30 years.

Laimonis Kavalieris

University of Otago

Counting breakpoints in time series

Structural breaks in time series include abrupt changes in mean, variance or autocorrelation either singly or in any combination. They have attracted a great deal of attention in the econometrics literature, for example time series with level shifts are used to model apparent long memory. Other recent work uses breakpoint models as approximations for non-stationary time series. We are concerned with the estimation of the number of structural breaks.

In this talk we discuss the Minimum Description Length (MDL) principle as a criterion for selecting the number of breaks in a level shift model in a positively correlated time series. Some new theoretical results will be mentioned together with links to smoothing time series with a drifting mean. The procedure will be illustrated using examples from the natural sciences.

Reuben Kendall and Mike Camden

Statistics New Zealand

Confidentiality of longitudinal collections

Longitudinal collections present information from the same 'reporting units' over multiple time periods. Such collections are increasing in importance due to their ability to reveal the dynamics of a population: transitions in variables for individuals can be related to transitions in other variables. This contrasts with cross-sectional collections which provide a 'snapshot' in time. However, the positive aspects provided by longitudinal data may pose increased risks of disclosure, with information showing transitions over time potentially being highly identifying of individuals.

Longitudinal analysis is relatively new in national statistical agencies. As a result, confidentiality procedures surrounding it are still being investigated and developed. This paper presents a discussion of the risks for providing access to aggregate data and microdata, the current confidentiality measures implemented by national statistical agencies, and the possible solutions for these confidentiality risks

Claudia Kirch

University of Karlsruhe, Germany

Bootstrapping sequential tests for change-point analysis

Change-point analysis studies whether the underlying model of an observed stochastic process changes at least once during the observational period. In a sequential setting we test the null hypothesis repeatedly after each new observation until a change is detected. This is specifically useful if data is collected automatically.

Critical values are usually based on distributional asymptotics, where the limit is taken with respect to a historic data set, which is used for model building. If this set is rather small the asymptotic distribution is not a good approximation.

Therefore we are interested in bootstrapping methods, which have been used extensively in classical testing but hardly at all in sequential situations. In such a setting we can make use of the new incoming observations for the bootstrap. From a practical point of view this is computationally expensive, so one needs some variations. From a theoretical point of view this means that we have new critical values with each incoming observation, so the question is whether this procedure remains consistent.

Frances Krsinich

Statistics New Zealand

Using hedonic regression to assess the housing rentals component of the New Zealand Consumers Price Index

The housing rentals index is the highest weighted component of the Consumers Price Index, contributing nearly 8 percent as at the June 2008 quarter. The estimation of non-Housing New Zealand rents is based on data from an area-based sample survey, updated with new bonds lodged with the Department of Building and Housing data. A 'matched sample' approach is used to remove the effect of changing sample composition, and therefore changing quality characteristics, from the measurement of price change. Although matched sample approaches for price change measurement are common across official statistics agencies, there is a potential risk that some pure price change implied by the changing population is removed.

As part of a more general review of the housing rentals index estimation, we used hedonic regression models to examine whether there is any residual pure price change being smoothed out of the index due to the matched sample approach. The longitudinal nature of the data enabled us to use fixed effects regression to control for all characteristics (observed and unobserved) fixed across time at the rental dwelling level.

Kuldeep Kumar

Bond University

Bankruptcy prediction using decision tree technique

Accurate business failure prediction models would be extremely valuable to many industry sectors, particularly in financial investment and lending. The potential value of such models has been recently emphasised by the extremely costly failure of high profile businesses in both Australia and overseas. In this paper, the various statistical techniques used in previous studies are presented and reviewed, including two alternative techniques that have produced promising results, namely survival analysis and decision trees.

Ivy Liu

Victoria University of Wellington

A hybrid method for longitudinal data

This is joint work with Richard Arnold (VUW), Fiona Gunasekara and Ken Richardson (Otago University)

This talk discusses a hybrid method that decomposes the time-varying predictors into two parts, one representing within-subject variation, the other representing between-subject variation. The method combines some of the features from fixed effects and random effects models. We consider various cases that model continuous, binary and ordinal responses. We also show a way to fit the proportional odds model, which is currently the most popular model for ordinal responses, using the conditional approach.

Magnus McGee

University of Otago, Christchurch

Bayesian Calibration in Microsimulation of Disease and Cancer in New Zealand (MoDCoNZ)

Colorectal cancer (CRC) is one of the major causes of illness and death in New Zealand (NZ). A microsimulation model to examine the effect of published guidelines and recommendations on treatment of CRC is presented. The model simulates a large number of individual life histories based on the New Zealand population and the epidemiology of CRC. Methods used by other modelers are briefly reviewed including the choice of parameterization and optimization methods. Bayesian calibration based around application of the Metropolis-Hastings algorithm and rejection sample methods are presented for the MoDCoNZ models.

Esther Meenken

Plant & Food Research

From town to farm: Compost in Canterbury agricultural systems

This is joint work with C. S. Tregurtha and A. J. Horrocks, both PFR

New green waste collection regimes are being introduced in an increasing number of New Zealand cities and towns in an effort to divert the amount of organic waste heading to landfills each year. These regimes include specialised collection of green waste, which is then turned into a nutrient rich compost.

The Canterbury agricultural sector could benefit from utilising compost produced within the region. Our initial research assessed the effects of compost on the establishment and growth of different forage crops, and the associated soil quality changes under each crop. Statistical issues included talking with multiple stakeholders about running a scientifically robust trial, designing trials on a site with trends in two directions, analysis of the data using a mixed model, and reporting on results.

Subsequent years will continue these on-farm trials, and extend the scope of the project to include an intensive arable cropping trial, several demonstration sites across Canterbury and a lab-based nutrient release experiment. The arable cropping trial presents a challenge as it will be set up in a location where a previous trial is likely to have influenced the uniformity of the site.

Rahul Mukerjee

Indian Institute of Management Calcutta

Highest posterior density regions based on empirical-type likelihoods: Role of data-dependent priors

We consider the Bayesian versus frequentist interface with reference to a very general class of empirical-type likelihoods which includes the usual empirical likelihood and all its major variants proposed in the literature. Probability matching priors play a key role in this context. It is known that none of these likelihoods admits a data-free probability matching prior for the highest posterior density region. We show that at least for the usual empirical likelihood this problem can be resolved if data-dependent priors are entertained. Necessary higher order asymptotics are developed for this purpose. The theoretical results are supported by a simulation study.

Richard Penny

Statistics New Zealand

Imputation variance estimation for the Accommodation Occupancy Survey

This is joint work with Raazesh Sainudiin, Department of Mathematics and Statistics, University of Canterbury

We formulate the problem of imputation variance estimation for the Accommodation Occupancy Survey (AOS) run on behalf of the Ministry of Tourism by Statistics New Zealand and develop a methodology to address this problem. We use Bayesian nonparametric methods and bootstrap techniques to provide consistent estimates of the imputation variance under the assumption of homogeneity within the predefined imputation cells.

Steve Quinn

Menzies Research Institute, University of Tasmania

A comparison of logistic regression goodness-of-fit statistics when applied to the log link

Blizzard L & Hosmer DW (2006, Biometrical Journal, pp.5-22) contributed to the development of the log binomial regression model for binary outcomes that make it possible to directly estimate relative risk in follow-up studies, and prevalence ratios in cross-sectional studies, and to adjust for continuous covariates. There are several goodness-of-fit statistics that describe how well a logistic regression model fits a set of observations, but to date little work has been done that summarizes the discrepancy between observed and expected values for log link models. Through extensive simulations we compare the performance of several goodness-of-fit statistics via rejection rates, power to detect an incorrectly specified model, and power to detect an incorrectly specified link when applied to log link models.

Alastair Scott

University of Auckland

Non-response in case-control studies

We develop the properties of a broad class of estimating equations designed to handle non-response in population-based case-control studies. The class includes survey-weighting along with a number of other methods some of which attain full semi-parametric efficiency. We compare the efficiency of the methods in simulations based on the Women's Cardiovascular Health Study (Schwartz et al 1977, Ann. Internal Medicine 127, 596-603).

Evan Stubbs

Solution Manager, Analytics – SAS ANZ

Feedback from the field: How our graduates are letting down the profession

With Ian Ayers, Malcolm Gladwell, and Thomas Davenport popularising analytics, it's not surprising to see an increased awareness of and desire for the power of statistics and modeling within the private and public sectors. Unfortunately, irrespective of the tools and techniques used, there exists a significant mismatch between the skills increasingly required by organisations and the abilities held by recent graduates. In this presentation, we look into why this is (with specific reference to a number of ANZ-based examples) and what must be done about it if an education in analytics is to remain relevant, useful, and robust.

Finlay Thompson

Dragonfly

Building a two stage Bayesian dolphin capture model with JAGS

This is joint work with Edward Abraham and Megan Oliver at Dragonfly, carried out on behalf of the Ministry of Fisheries

Common dolphin are killed by trawlers targeting jack mackerel off the west coast of the North Island. Estimates of the total number of dolphins captured are made using a two stage Bayesian capture model, fitted using Markov chain Monte Carlo techniques. About 20% of the jack mackerel trawl fishery has been observed, and the model is used to estimate captures on the unobserved trawls.

The model estimates the probability of a capture event occurring (first stage) using a hierarchical general linear model. The number of dolphins caught (stage two), is then estimated from a zero-truncated Poisson distribution. This structure reflects the pragmatic need to deal with data that contains a lot of zeros, and the behavioural process involved: dolphins tend to swim together, and get caught together.

The model is specified using the BUGS language, and fitted using JAGS (Just Another Gibbs Sampler), a program developed by Martyn Plummer. A fully scripted process was developed for this project, taking data from extracts, through grooming, modelling, plotting, to final publication. A description of this process stack is given, along with advantages and problems identified along the way. A range of open source tools was used, including Postgresql, R, JAGS, Sweave, Latex, all glued together with Git, Bash, Make and Linux.

Peter Thomson

SRA

Mixed methods for fitting the GEV distribution

The generalised extreme-value (GEV) distribution is widely used for modelling and characterising extremes. It is a flexible 3-parameter distribution that combines three extreme-value distributions within a single framework: the Gumbel, Frechet and Weibull. Common methods used for estimating the GEV parameters are the method of maximum likelihood and the method of L-moments.

This paper generalises the mixed maximum likelihood and L-moments GEV estimation procedures proposed by Morrison and Smith (2002) and derives the asymptotic properties of the resulting estimates. Analytic expressions are given for the asymptotic covariance matrices in a number of important cases, including the estimators proposed by Morrison and Smith (2002). These expressions are verified by simulation and the efficiencies of the various estimators established.

The asymptotic results are compared to those obtained for small samples, and the properties of the various estimators, including constrained maximum likelihood estimators, are considered. The corresponding quantile estimators are also assessed for accuracy and bias. Using simplified constraints for the support of the log-likelihood, computational strategies and graphical tools are developed which lead to computationally efficient, numerically robust, estimation procedures. These methods are also applied to 24-hour annual maximum rainfall at Wellington, New Zealand, over the period 1940-1949 and within each phase of the Interdecadal Pacific Oscillation (IPO).

Rolf Turner

Starpath Project, University of Auckland

A dearth of evidence

These days it is very trendy for authorities to demand that initiatives and decisions be "evidence based". This is certainly as true in the education milieu as it is anywhere. The difficulty is that it is, in many cases, virtually impossible to produce "statistically significant" evidence of the success of an initiative. The initiative would have to produce an overwhelming change in order for the improvement to show up as being "significant". This is because of the level of variability with which one has to contend and of the ever-bedevilling fact that precision increases with the square root of the sample size and not with the sample size. I shall give some examples of this problem taken from work done in the context of the Starpath Project.

Chikako van Koten

AgResearch Limited

Longitudinal analysis of count data: A GEE approach

Analysis of count data collected from groups of samples over a period of time is one of the interesting problems in Statistics. I recently came across such data and analysed them using a GEE (Generalized Estimating Equations) approach. In this talk, the result from this GEE approach is presented and compared with results from other approaches.

Dong Wang

Victoria University of Wellington

Properties of a random matrix from a multivariate normal distribution

In this communication, we consider a p by n random matrix $X = (x\_1, ..., x\_n)$ from a pn-dimensional multivariate normal distribution, where x_i and x_j are correlated. The covariance matrix of x_i and x_j is given by the p by p matrix V_{ij} and each x_i is from a p-dimensional multivariate normal distribution with a mean vector and a covariance matrix which is given by the Kronecker product. The statistical objective is to consider the maximum likelihood estimate of the mean matrix and various components of the covariance matrix and their statistical applications. Some properties of these estimators are also investigated in this paper.

Norio Watanabe

Chuo University, JAPAN

Estimation of average trajectory of nearly periodic motion

In this study we discuss an estimation problem for the data obtained from almost periodic movement. For example, successive strokes of the swimmer or location of some body part of a walking man can provide such data by setting the center of the body. Similar data are found in the sports analysis or health science. Usually data is obtained through motion capture and can be converted to 2-dimensional time series. A purpose is to find the average movement. A usual approach for estimation of the average trajectory is based on the functional data analysis. In this approach whole time series is devided into piecewise series according to each periods and regarded as the functional data. And the average trajectory can be estimated as the mean of the functional data. For successive movement, however, the obtained trajectory is not necessarily closed. Moreover the registration is a bothering problem. Our approach is based on the time series analysis. First we fit some nonlinear time series model to the data. Then we can estimate the average trajectory as an attractor of a nonlinear system by using the estimated model. In this study we use a multilayered neural network as a nonlinear time series model and discuss applicability of our approach by simulation studies.

Robin Willink

Industrial Research Ltd

A general construction of a shrinkage confidence interval

This paper describes the construction of confidence interval that is narrower than a standard interval when a prior guess, theta_0, for an unknown univariate parameter theta is sufficiently accurate. This confidence interval is obtained through moving one or both of the limits of a standard interval toward theta_0. A general procedure for the construction of such a `shrinkage confidence interval' is given for suitable univariate problems. When theta_0 is close to theta, the technique can achieve an interval with mean width considerably smaller than that of the corresponding standard interval.

This paper builds on work described in `Shrinkage confidence intervals for the normal mean: using a guess for greater efficiency', R. Willink, Canadian Journal of Statistics, 36, No 4, (2008) 623-638. (Note: This form of shrinkage estimation is to be distinguished from James-Stein estimation.)

Kit Withers

Industrial Research Ltd

The distribution of the maximum of a first order moving average: The continuous case

We give the distribution of M_n, the maximum of a sequence of n observations from a moving average of order 1. Solutions are first given in terms of repeated integrals and then for the case where the underlying independent random variables have an absolutely continuous density. When the correlation is positive,
$$
P(M_n \max^n_{i=1} X_i \leq x)\ =\ \sum_{j=1}^\infty \beta_{jx} \ \nu_{jx}^{n} \ \approx \ B_{x}\ \nu_{1x}^{n}
$$
where $\{X_i\}$ is a moving average of order 1 with positive correlation, and $\{\nu_{jx}\}$ are the eigenvalues (singular values) of a Fredholm kernel and ν_{1x} is the eigenvalue of maximum magnitude. A similar result is given when the correlation is negative. The result is analogous to large deviations expansions for estimates, since the maximum need not be standardized to have a limit.

For the continuous case the integral equations for the left and right eigenfunctions are converted to first order linear differential equations. The eigenvalues satisfy an equation of the form
$$
\sum_{i=1}^\infty w_i(\lambda-\theta_i)^{-1}=\lambda-\theta_0
$$
for certain known weights $\{ w_i\}$ and eigenvalues $\{ \theta_i\}$ of a given matrix. This can be solved by truncating the sum to an increasing number of terms.

Lingyun Zhang

Massey University, Wellington

A note on Bartlett's M test for homogeneity

After pointing out a drawback in Bartlett's chi-square approximation, we suggest a simple modification and a Gamma approximation to improve Bartlett's M test for homogeneity of variances.

Guan Yu Zheng

Statistics New Zealand

Viability of reducing survey frequency using administrative data

Administrative data is data collected to fulfill administrative purposes, instead of statistical purposes. Statistics New Zealand would like to expand the use of administrative data in various ways. One way is to investigate the possibility of reducing survey frequency by using administrative data in time series. Using administrative data to reduce survey frequency requires that administrative data possess highly similar statistical properties to survey data. In this paper, we explore whether Electronic Card Transactions (ECT) as an administrative data source is a good leading indicator of the Retail Trade Survey (RTS), as a first step towards assessing the viability of reducing survey frequency. For Statistics New Zealand, a good leading indicator should provide high correlations and similar seasonal adjustment estimates to the actual ones. We found ECT and RTS are highly cointegrated with constant seasonal variation. In one step ahead forecasting, forecasts are unbiased and the differences in seasonal adjustments are relatively small for subgroups of RTS.

NZSA 2009 Abstracts for Poster Presentations

Abstracts appear alphabetically, ordered by the presenter's last name

Simon Anastasiadis

Victoria University of Wellington

Understanding vehicle warranty claims from the observed driving pattern

Warranty claims arise as a result of product failure; the better manufacturers understand claims the more accurately they can plan to meet the costs of those claims and the clearer the quality of their product becomes. Vehicle failures depend on the usage of each vehicle. While manufacturers can track the age of the vehicle they can not observe the usage it is subject to, except at random points in time; i.e., when warranty claims are made.

We wish to infer the actual vehicle usage from the observed usage when warranty claims are made. An attempt is made to find a relationship between vehicle warranty claims and the driving pattern that can be observed at the time of warranty claims.

Giorgi Kvizhinadze and Haizhen Wu

Victoria University of Wellington

Diversity analysis in multiple choice questionnaires

Consider a multiple choice questionnaire with q questions, and n individuals are asked to fill out this questionnaire so we obtain n "opinions". Our questions are: how many different opinions will we observe? What is the proportion of unique opinions? What is the proportion of opinions we will see twice or any k times?

One would reasonably assume that the answers to our questions depend on various properties of the questionnaire structure, such as the number of questions, the number of possible answers in each question, the probabilities of the answers, the number of interviewees and so on. However, we discovered that the asymptotic behaviour of the quantities we are interested in follows the Karlin-Rouault law, which depends on only one parameter that incorporates all information about the properties of the questionnaire structure.

Xiaomei Li

Victoria University of Wellington

Computational model selection for the Poisson regression model

The task of statistical model selection is to choose a parsimonious model from a collection of models, which gives the best approximation to the observed data. The ultimate objective for this project is to illustrate the non-Bayesian and Bayesian approaches in computational model selection for the Poisson regression model. We developed three commonly used model selection methods by using R and WinBUGS. They are developed based on hypothesis testing, deviance, Bayesian approach, and information criterion. Also, we compare the results which we obtain from the different approaches.

Shaochuan Lu

Victoria University of Wellington

Occurrence patterns of New Zealand deep earthquakes: Characterized by MMPP and its extensions

In this poster, we suggest the main occurrence pattern of the New Zealand deep earthquakes, in a relatively large time scale, is the time-varying seismic activities. The mechanisms and reasons behind this time-varying behavior are still not well understood. We propose a switching Poisson model, which is a Poisson process with the Poisson rates switching between two levels according to an unobserved two-state Markov chain, to characterize the time-varying seismic activities.

Her Guan Teo

Victoria University of Wellington

Nonparametric estimation of the mean cumulative number/cost of automotive warranty claims

Automotive warranty involves both age and mileage limits. Age is known for all sold vehicles but the accumulated mileage is usually not known. The method introduced here uses a simple linear approach in modeling accumulated mileage, and the mean cumulative number/cost of automotive warranty claims is estimated as a function of age, mileage, as well as the actual time.

David Walsh

Victoria University of Wellington

Bayesian methods of earthquake focal mechanism estimation
This is joint work with Richard Arnold and John Townend, both VUW

Earthquakes are important signatures of the stress in the earth's crust. In order to interpret the constraints earthquakes place on the stress field it is important to summarise — in a statistically valid way — the information that observations of earthquakes contain. The focal mechanism of an earthquake describes the geometry of the fault on which the earthquake occurred using three parameters: the strike, dip and rake. We are investigating a Bayesian method of calculating a posterior probability distribution for these focal mechanism parameters. We approximate such a distribution using a Matrix-Fisher distribution, and here we apply these methods to phase data from the Raukumara Peninsula, New Zealand. Our approach has the advantage of enabling us to incorporate observational errors and give robust posterior uncertainties for the estimated focal mechanism parameters.

Robin Willink

Industrial Research Ltd

Mathematical entropy and a criticism of a usage of maximum-entropy distributions

The 'principle of maximum entropy' is sometimes used for assigning prior distributions to unknown parameters in Bayesian analyses. Similarly, it has been advocated for choosing a probability distribution to represent individual unknowns such as systematic deviations in measurement problems (Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method, JCGM 101:2008, [available from http://www.bipm.org/en/publications/guides/gum.html].) We show how supporting claims like 'the maximum-entropy distribution is minimally committal' are indefensible in this context. We examine the origin of the idea of entropy in communication theory, and conclude that the idea that entropy measures 'information' only has meaning with sequences of categorical random variables. So it cannot legitimately be associated with 'information' about individual parameters.

Lisa Woods

Victoria University of Wellington

A probabilistic method of tectonic stress estimation

Earthquakes are the result of tectonic stress, which builds up along faults and plate boundaries. There is no direct method capable of measuring tectonic stress within the earth's lithosphere so we record earthquakes - an effect of the stress which we are able to observe. The stress inversion process that we discuss here uses earthquake observations to determine the compressive stresses acting within a region of the earth by applying a Bayesian approach, so we are using a probabilistic method of tectonic stress estimation.

Estimating tectonic stress has many useful applications such as the possible prediction of volcanic eruptions because even though most tectonic stress is the result of plate movement, volcanic systems can cause extra stress to build up in surrounding areas due to the movement of magma within the system (Twiss & Moores 2007).

Australian and New Zealand Journal of Statistics: Present and future?

A plenary session on the future of the Australian and New Zealand Journal of Statistics will run from **1.00-2.00pm on Thursday 3 September 2009 in Cotton Lecture Theatre LT122, Victoria University of Wellington**. The session will be **chaired by Ian Westbrooke** (Department of Conservation, Christchurch).

There will be four speakers (three 12-15 minute slots for Steve, Sue and Murray and one five minute slot for Roger), in the following order.

Steve Haslett (Massey University, Palmerton North), the **ANZJS Managing Editor**:
- a focus on the current situation with ANZJS and plans for the future content of the journal.

Roger Littlejohn (AgResearch, Invermay Agricultural Centre), **ANZJS Management Committee member**:
- a five minute slot for a brief update on ANZJS finances.

Sue Wilkins (RSNZ), **Publishing Operations Manager at the Royal Society of New Zealand**:
- an overview of the changing journal environment internationally, and particularly for journals based in NZ and Australia
- briefly outline RSNZ plans for the journals it publishes
- thoughts on the issues facing constituent RSNZ societies like NZSA that publish journals (jointly with SSAI in our case).

Murray Jorgensen (University of Waikato), **current ANZJS Management Committee member and previously Editor of the New Zealand Statistician, then Applications Editor of ANZJS**:
- an outline of personal views on some options for future publishing models for the ANZJS.

The plan is to take short questions immediately after each speaker, and reserve time for broader discussion at the end.

A report on the session will be published in the **NZSA newsletter**, following the conference.

Statistical Education Workshops and Talks

Mathematics and Statistics teachers are welcomed to the New Zealand Statistical Association 2009 Conference on Thursday 3 September for an afternoon of workshops and talks on statistical education.

The Focus of the Workshops

The new *NZ Curriculum* contains the learning area *Mathematics and Statistics*, but it also contains *Principles, Values* and *Key Competencies*. The *Statistics* strand of our learning area involves real-life context, and so is rich in opportunities for students to practice the *Key Competencies*, and to have experiences that embody the *Principles* and *Values*. In each workshop, we will explore these emphases in the context of teaching and learning about statistics.

1.00-2.00 pm: Two sets of parallel workshops for teachers, in CO 118 and CO 119: hands-on activities for use in the classroom, plus demonstrations of web resources and/or use of computers.

Time	Workshop	Location
1.00-1.30pm	Maxine Pfannkuch: *Comparative statistical reasoning: Some essential dialogues*	CO 118
1.00-1.30pm	Tim Burgess: *The challenges of using statistical investigations in primary school classrooms*	CO 119
1.30-2.00pm	Paul Bucknall, Mike Camden, Nathaniel Pihama: *A website of statistical riches*	CO 118
1.30-2.00pm	Derek Smith: *Probability trees and two way tables*	CO 119

2.00-4.00 pm: Five 20-minute presentations in CO LT122, and an **afternoon-tea break in Maclaurin building Foyer**; **Chair: Mike Camden**. Because of the popularity of the education sessions at previous NZSA conferences, this will be the only stream of presentations running at this time.

Time	Talk	Location
2.00-2.20pm	Mike Camden: *New resources for learning of the Curriculum's statistics*	CO LT122
2.20-2.40pm	John Harraway: *A new teaching resource for Year 13*	CO LT122
2.40-3.00pm	Kevin Burns: *How smart are birdbrains? Number sense in the New Zealand robin*	CO LT122
3.00-3.20pm	**Afternoon Tea/Coffee**	Maclaurin Foyer
3.20-3.40pm	Robin Averill: *Learning in Statistics: Perspectives of secondary school students*	CO LT122
3.40-4.00pm	Alasdair Noble and Anne Lawrence: *One good tern...*	CO LT122

4.00-5.00 pm: Two parallel workshops for teachers, in CO 118 and CO 119.

Time	Workshop	Location
4.00-5.00pm	Robin Averill: *Ideas for teaching statistics in junior secondary school classrooms*	CO 118
4.00-5.00pm	Sandi Tait-McCutcheon: *Framing Statistics and Probability Level 1 to 5*	CO 119

NZSA 2009 Abstracts for Statistical Education Talks

Abstracts appear alphabetically, ordered by the speaker's last name

Robin Averill

Victoria University of Wellington

Learning in Statistics: Perspectives of secondary school students

Quality learning tasks are, on their own, insufficient to engender effective learning. This talk will present a range of perspectives of secondary school students into characteristics of teacher student relationships they find conducive to learning statistics. Dispositional and culturally responsive themes will be discussed, supported by qualitative and quantitative data drawn from a recent study across six multicultural classes from three schools. Findings indicate that teachers who know their students as individuals and as learners, acknowledge students' cultures, and address issues of respect and tolerance whilst holding high expectations of achievement are well placed to maximise students' learning.

Kevin Burns

Biological Sciences, Victoria University of Wellington

How smart are birdbrains? Number sense in the New Zealand robin
This is joint work with Simon Hunt and Jason Low, both VUW

Knowledge of mathematics is thought to distinguish humans from other animals. However, the potential for wild animals to use number discriminations to solve ecological problems is poorly resolved. We conducted a series of experiments to test whether a food-hoarding songbird, the New Zealand robin Petrocia australis, could discriminate between cache sites containing different numbers of prey. We found that robins utilize a highly sophisticated number sense to retrieve and pilfer stored food, thus providing a critical link in our understanding of the adaptive significance of numerical competency in untrained wild animals.

Mike Camden

Statistics New Zealand

New resources for learning of the Curriculum's statistics

The real-life nature of statistics means that we're well-placed to use novel communication methods in the learning of our discipline. We'll look at The NZ Curriculum's 'capabilities for living and lifelong learning', then experience some new resources, including LearningMedia's new books Statistics in the Media, and some interactive and dynamic data products on Statistics New Zealand's website.

John Harraway

University of Otago

A new teaching resource for Year 13

To be presented by Mike Camden (Statistics New Zealand and the NZSA Education Committee)

A second set of 15 edited DVD presentations on research projects using statistics has been completed. Associated MS Excel data files contain all data and classroom tasks are suggested. Sections of one DVD will be shown along with data and task details. Included is a resampling procedure for MS Excel which is easy to use without other MS Excel add ons. The 15 DVD's, which cover in total four hours of presentation, are about to be placed on the University of Otago unitube site. Another website with links to unitube is being established for data, associated tasks and the bootstrap procedure. Feedback from teachers is welcome. If worthwhile, more projects will be filmed in areas of interest to school students, like rugby injuries and youth motor accidents.

Alasdair Noble and Anne Lawrence

IFS/Statistics & Centre for Educational Development, Massey University

One good tern...

A data set will be described and then a series of activities relevant to various age groups and topics with the Mathematics and Statistics curriculum will be introduced. A number of worksheets applicable to different parts of the data set will be distributed.

NZSA 2009 Abstracts for Statistical Education Workshops

Abstracts appear alphabetically, ordered by the speaker's last name

Robin Averill

Victoria University of Wellington

Ideas for teaching statistics in junior secondary school classrooms

This interactive session will present several rich statistical activities suitable for teaching level 4-6 statistics. The activities will be discussed in terms of how they can be used to develop statistical learning while promoting the curriculum principles and values relating to setting high expectations, being responsive to the principles of the Treaty of Waitangi and to cultural diversity, and providing opportunities for classroom members to participate, contribute, and value and develop their relationships with others.

Paul Bucknall, Mike Camden and Nathaniel Pihama

Statistics New Zealand

A website of statistical riches

Our new website is a treasure chest of riches for school statistics. We put many of the treasures there, not specifically for schools, but for our public audience. This means that they have daily-life contexts. We'll show how teachers can use them not only to reach statistical achievement outcomes, but also to develop the Curriculum's 'capabilities for living and lifelong learning'.

Tim Burgess

Massey University

The challenges of using statistical investigations in primary school classrooms

We will consider some of the challenges with statistical thinking and reasoning in primary school classrooms and the links to key competencies. Use of censusatschool will also be looked at.

Maxine Pfannkuch

The University of Auckland

Comparative statistical reasoning: Some essential dialogues

In this session we will discuss data dialogues that are essential for understanding and verbalizing the rich conceptual repertoire inherent in the comparison of box plots and in making inferences.

Derek Smith

New Zealand Association of Mathematics Teachers & Wellington Mathematics Association

Probability trees and two way tables

Probability trees are more widely used than two way tables in mathematics and statistics classrooms. What are some of the issues that students have to grapple with and should teachers be addressing multiple representations of probability rather than defaulting to a probability tree model?

Sandi Tait-McCutcheon

Victoria University of Wellington

Framing Statistics and Probability Level 1 to 5

A team of 14 lead teachers from the Parumoana area in Wellington have designed a draft framework for teaching and learning Statistics and Probability Level 1 to 5. During this interactive session participants will engage with a selection of rich statistical activities that complement our framework. Lead Teachers will share their practice and knowledge.

NZSA 2009 Conference and Semiparametric Regression Workshop: Participants

Last name	First name	Email	Affiliation
Abraham	Edward	(all removed for version on the web)	Dragonfly, Wellington
Aidara	Cherif		Victoria University of Wellington
Aldridge	Tony		Christchurch
Anastasiadis	Simon		Victoria University of Wellington
Anwar	Muhammed Nafees		Massey University, Palmerston North
Archibald	Judith		Statistics New Zealand, Wellington
Arnold	Richard		Victoria University of Wellington
Ball	Rod		Scion, Rotorua
Ball	Tim		Tim Ball Limited
Barton	James		James P Barton & Associates Limited
Bascand	Geoff		Statistics New Zealand, Wellington
Betz-Stablein	Brigid		Massey University, Palmerston North
Bhattacharya	Aloka		Victoria University of Wellington
Bilton	Penny		Massey University, Palmerston North
Bracewell	Paul		Offlode, Wellington
Brown	Jennifer		University of Canterbury
Brunning	Deborah		Statistics New Zealand, Wellington
Bryant	John		Statistics New Zealand, Christchurch
Camden	Mike		Statistics New Zealand, Wellington
Castillo	Elenita		Ministry of Education, Wellington
Chang	Keng Hao (Danny)		University of Auckland
Chattopadhyay	Saibal		Indian Institute of Management Calcutta
Chee	Chew-Seng		University of Auckland
Chukova	Stefanka		Victoria University of Wellington
Clark	Austina		University of Otago
Clark	Megan		Victoria University of Wellington
Cowie	Paul		IAG New Zealand Limited, Wellington
David	Irene		University of Canterbury
Davies	Robert		Statistics Research Associates, Wellington
Davies	Tilman		Massey University, Palmerston North
Davis	Walter		Statistics New Zealand, Wellington
Dawber	James		University of Canterbury
Degnan	James		University of Canterbury
Devlin	Miranda		Ministry of Health
Doherty	Mike		Statistics New Zealand, Wellington
Doonan	Ian		NIWA, Wellington
Easton	Brian		Wellington
Enright	Jamas		The Treasury, Wellington
Ferguson	Andrew		Plant & Food Research, Te Puke
Fernando	Sarojinie		Massey University, Palmerston North
Forbes	Sharleen		Statistics New Zealand, Wellington
Fu	Dan		NIWA, Wellington
Ganesalingam	Selvanayagam		Massey University, Palmerston North
Gee	Graeme		Offlode, Wellington
Grant	William		IRD, Wellington
Green	Peter		University of Otago
Harte	David		Statistics Research Associates, Wellington
Haslett	Stephen		Massey University, Palmerston North
Hawkes	Timothy		Statistics New Zealand, Wellington
Haywood	John		Victoria University of Wellington
Hazelton	Martin		Massey University, Palmerston North
Henderson	Harold		AgResearch Ruakura
Hindmarsh	Diane		University of Wollongong, Australia
Hirose	Yuichi		Victoria University of Wellington
Hodge	Miriam		University of Canterbury
Isidro	Marissa		Massey University, Palmerston North
Jackson	Fraser		Wellington
Johnston	Steven		Ministry of Social Development, Wellington
Jorgensen	Murray		University of Waikato
Jowett	Clint		Offlode, Wellington
Kavalieris	Laimonis		University of Otago
Kendall	Reuben		Statistics New Zealand, Wellington
Kirch	Claudia		University of Karlsruhe, Germany
Koolaard	John		AgResearch Palmerston North
Kozik	Hinrich		Victoria University of Wellington
Krsinich	Frances		Statistics New Zealand, Wellington
Kumar	Kuldeep		Bond University, Australia
Kuzmicich	Stephen		Wellington
Lankshear	Lisa		Victoria University of Wellington
Lee	Alan		University of Auckland
Li	Xiaomei		Victoria University of Wellington
Littlejohn	Roger		AgResearch Invermay
Liu	Jiale (Carlos)		Statistics New Zealand, Wellington
Liu	Ivy		Victoria University of Wellington

Last name	First name	Email	Affiliation
Lu	Shaochuan	(all removed for version on the web)	Victoria University of Wellington
Luo	Dongwen		AgResearch Palmerston North
Luo	Jing (Annie)		AMI, Christchurch
MacDonald	Anna		University of Canterbury
Mansour	Aghababaei Jazi		Massey University, Palmerston North
Mao	Tian		Massey University, Palmerston North
McDonald	Barry		Massey University Albany
McGee	Magnus		Christchurch School of Medicine
McGinty	Peter		Statistics New Zealand, Wellington
McWhirter	Judi		University of Waikato
Meade	Aleisha		Massey University, Palmerston North
Meenken	Esther		Plant & Food Research, Lincoln
Mormede	Sophie		NIWA, Wellington
Mukerjee	Rahul		Indian Institute of Management Calcutta
Nair	Vijay		University of Michigan, USA
Obolonkin	Vladimir (Vlad)		Livestock Improvement Corporation, Hamilton
Olubode-Awosola	Femi		AgResearch Palmerston North
Park	Zaneta		AgResearch Palmerston North
Parry	Katharina		Massey University Albany
Penny	Richard		Statistics New Zealand, Christchurch
Pfannkuch	Maxine		University of Auckland
Pledger	Shirley		Victoria University of Wellington
Potter	Fred		AgResearch Palmerston North
Qiao	Chungui		Ministry of Social Development, Wellington
Quinn	Steve		University of Tasmania
Ramsay	Diane		Statistics New Zealand, Wellington
Rea	Alethea		University of Auckland
Reilly	James		Statistical Insights Ltd, Auckland
Rohan	Maheswaran		Department of Conservation, Hamilton
Rouhollahi	Sima		Victoria University of Wellington
Ryan	Michael		Statistics New Zealand, Christchurch
Scott	Alastair		University of Auckland
Sibanda	Nokuthaba		Victoria University of Wellington
Smith	Murray		NIWA, Wellington
Song	Soon		Ministry of Social Development, Wellington
Stubbs	Evan		SAS, Melbourne
Su	Kristy		Harmonic Ltd, Wellington
Teo	Her Guan		Victoria University of Wellington
Thompson	Finlay		Dragonfly, Wellington
Thomson	Peter		Statistics Research Associates, Wellington
Turner	Rolf		Starpath Project, University of Auckland
Upsdell	Martin		AgResearch Ruakura
van Koten	Chikako		AgResearch Lincoln
Vere-Jones	David		Victoria University of Wellington
Visch	Julian		Wellington
Wand	Matt		University of Wollongong, Australia
Wang	Yong		University of Auckland
Wang	Dong		Victoria University of Wellington
Wang	Ting		Massey University, Palmerston North
Watanabe	Norio		Chuo University, Japan
Westbrooke	Ian		Department of Conservation, Christchurch
Willink	Robin		Industrial Research Ltd, Lower Hutt
Withers	Kit		Industrial Research Ltd, Lower Hutt
Wood	Michelle		IAG, Auckland
Woods	Lisa		Victoria University of Wellington
Wrathall	Wa		Massey University, Palmerston North
Wu	Shirley		Harmonic Ltd, Wellington
Wu	Haizhen		Victoria University of Wellington
Yeh	Li-Chia		Ministry of Health
You	Ryan		SHORE, Massey Univeristy
Young	Jim		Christchurch
Zhang	Lingyun (Larry)		Massey University Wellington
Zheng	GuanYu (Fish)		Statistics New Zealand, Christchurch

NZSA 2009 Education Session: Additional Participants

Last name	First name	Email	Affiliation
Averill	Robin	(all removed for version on the web)	SEPI, Victoria University of Wellington
Bucknall	Paul		Statistics NZ
Burgess	Tim		Massey University
Burns	Kevin		Victoria University
Devlin	Raewyn		Tokomairiro High School
Green	Colin		Wellington College
Haley	Siobhan		Wellington Girls' College
Harvey	Darrell		Wellington College
Harvey	Roger		Victoria University
Hawke	Aaron		Victoria University
Izaz	Charmaine		Wellington College
Lawrence	Anne		Massey University College of Education
Noble	Alasdair		Massey University
Pihama	Nathaniel		Statistics New Zealand
Ross	Kelly-Ann		St Theresas, Plimmerton
Sharman	Jane		Christchurch Boys' High School
Smith	Derek		The Correspondence School
Sun	Jane		Wellington College
Tait-McCutcheon	Sandi		Victoria University of Wellington
Tideswell	Andrew		Samuel Marsden Collegiate
VanMilligan	Linda		Wellington Girls' College
Wilson	Dave		St Mary's College, Wellington