ANU-UC-ISM Joint Symposium on Environmental Statistics 2016

Venue: University of Canberra (University Drive, Bruce ACT 2617, Australia)

Date: 3 December 2016

ANU-UC-ISM Joint Symposium on Environmental Statistics 2016

The Mathematical Sciences Institute of the Australian National University and the Institute of Statistical Mathematics have an agreement to undertake academic and research collaboration to develop mutually beneficial, creative, and productive scholarly activities in the field of statistics with applications in risk related areas. At this symposium, organised with the University of Canberra, we aim to promote the development of statistical methods for dealing with environmental problems.

Date

3 December 2016

The team behind ANU-UC-ISM Joint Symposium

Organizers:
Alan Welsh (ANU)
Shuangzhe Liu (UC)
Robert Clark (NIASRA&UoW)
Kunio Shimizu (ISM)
Koji Kanefuji (ISM)

Host Organization:
Mathematical Sciences Institute, The Australian National University, Australia
The Faculty of Education, Science, Technology and Mathematics, University of Canberra, Australia
The Institute of Statistical Mathematics, Japan

Program

09:00 | 09:10	**Opening address** *Alan Welsh (ANU)* **Welcome address** *Dharmendra Sharma (UC)*
Chair	**Shuangzhe Liu (UC)**
09:10 | 09:55	Statistical efficiency in distance sampling *Robert Clark (National Institute for Applied Statistics Research Australia & University of Wollongong)*
09:55 | 10:40	Accounting for misclassification errors in the estimation of transition probabilities: a spatial multistate dynamic site occupancy model *Reiichi Fukaya (The Institute of Statistical Mathematics)*
10:40 | 11:00	**Morning tea**
Chair	**Toshinari Kamakura (Chuo University)**
11:00 | 11:45	Multi-species distribution modeling using penalized mixture models *Francis Hui (The Australian National University)*
11:45 | 12:30	Under what conditions does hybridization promote adaptive radiation? *Kotaro Kagawa (The Institute of Statistical Mathematics)*
12:30 | 01:15	**Lunch**
Chair	**Robert Clark (NIASRA&UOW)**
01:15 | 02:00	Effective search for masked explanatory variables in linear regression *Yoshinori Kawasaki (The Institute of Statistical Mathematics)*
02:00 | 02:45	Where are they? – Using Bayesian approaches to estimate detection and occupancy rates of koalas *Bernd Gruber (University of Canberra)*
02:45 | 03:00	**Afternoon Tea**
Chair	**Koji Kanefuji (ISM)**
03:00 | 03:45	Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating *Tao Zou (The Australian National University)*
03:45 | 04:30	On circular correlation for data on the torus *Xiaoping Zhan (Southwestern University of Finance and Economics), Tiefeng Ma (Southwestern University of Finance and Economics), Shuangzhe Liu (University of Canberra), and Kunio Shimizu (The Institute of Statistical Mathematics)*
04:30 | 04:35	**Closing address** *Kunio Shimizu (ISM)*

Speakers

Robert Clark (NIASRA&UoW)
Francis Hui (ANU)
Tao Zou (ANU)
Bernd Gruber (UC)
Shuangzhe Liu (UC)
Yoshinori Kawasaki (ISM)
Keiichi Fukaya (ISM)
Kotaro Kagawa (ISM)

Robert Clark - National Institute for Applied Statistics Research Australia & University of Wollongong

Title Statistical efficiency in distance sampling

Abstract Distance sampling is a technique for estimating the abundance of animals or other objects in a region, allowing for imperfect detection. The detection rate is modelled as a function of distance depending on unknown parameters. Uncertainty about these parameters leads to an inflation in the variance of the abundance estimator. This inflation penalty is derived asymptotically, under the strong assumption that objects are uniformly distributed across the region. It is also obtained in a simulation study which generates non-uniform object locations using a Dirichlet process. The variance penalty is found to often be substantial, typically at least 2 but sometimes much higher, particularly for steeply declining detection rates. As a result, strip transects may out-perform distance sampling in some cases.

Francis Hui - The Australian National University

Title Multi-species distribution modeling using penalized mixture models

Abstract Multi-species distribution modeling, which relate the occurrence of multiple species to environmental covariates, is an important tool for both predicting how a species community responds to the environment, and identifying important covariates driving species co-occurrences. In this talk, I introduce a recently developed tool for modeling species communities called Species Archetype Models (SAMs). SAMs use a finite mixture of regression model to cluster species with similar environmental responses to form "archetypal response groups", and models these archetypal responses using a GLM, for instance.

A challenging part of using SAMs are that finite mixture of regression models is variable selection, given the large number of possible candidate models. I consider using penalized likelihood methods to accomplish this, and propose two penalties called MIXGL2 and MIXGL1 which exploit the inherent group structure of the coefficients in SAMs. In doing so, both penalties are able to simultaneously remove a covariate from all components (archetypes). Selection consistency and oracle properties associated with the two penalties are shown, with simulations suggesting MIXGL2 and MIXGL1 outperform other penalties that do not take into account the group structure of the coefficients. The application of penalized SAMs is illustrated on a multi-species dataset collected from the Great Barrier Reef off the north-east coast of Australia.

Tao Zou - The Australian National University

Title Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating

Abstract By learning the PM2.5 readings and meteorological records from 2010-2015, the severity of PM2.5pollution in Beijing is quantified with a set of statistical measures. As PM2.5 concentration is highly influenced by meteorological conditions, we propose a statistical approach to adjust PM2.5concentration with respect to meteorological conditions, which can be used to monitor PM2.5pollution in a location. The adjusted monthly averages and percentiles are employed to test if the PM2.5 levels in Beijing have been lowered since China's State Council set up a pollution reduction target. The results of the testing reveal significant increases, rather than decreases, in the PM2.5concentrations in the years 2013 and 2014 as compared with those in year 2012. We conduct analyses on two quasi-experiments—the Asia-Pacific Economic Cooperation meeting in November 2014 and the annual winter heating—to gain insight into the impacts of emissions on PM2.5. The analyses lead to a conclusion that a fundamental shift from mainly coal-based energy consumption to much greener alternatives in Beijing and the surrounding North China Plain is the key to solving the PM2.5 problem in Beijing.

Bernd Gruber - University of Canberra

Title Where are they? – Using Bayesian approaches to estimate detection and occupancy rates of koalas

Abstract Occupancy rates of animals populations are an often requested quantity in conservation ecology to monitor population trends. Occupancy rates can be estimated using occupancy models. To estimate parameters that affect detection and occupancy, Bayesian approaches are becoming more and more popular. Bayesian occupancy models are hierarchical models that allow to link occupancy data to ecological predictors and the detection process. We demonstrate the implementation of such an approach to estimate detection probability and occupancy rates of koalas to inform the establishment of a koala monitoring programme.

In a first experiment we estimated detection rates of koalas. We found detection probability was depended on the number of koala pellets under a tree and on the individual observer. We further explored the suitability to use time to detection within a hierarchical Bayesian framework.

Knowledge on the detection probability in turn was used to estimate the power to monitor population trends of koalas depending on sample size and spatial design. We demonstrate the advantage of the Bayesian approach as it seamlessly allows to incorporate the estimation of derived parameters such as the probability to detect a negative (positive) population trend. This kind of type-II errors and accompanying confidence intervals are much harder to estimate via frequentist approaches.

In addition the ability to calculate derived parameters is of importance to communicate findings of the approach to stakeholders involved in the design and planning of monitoring programmes.

Xiaoping Zhan - Southwestern University of Finance and Economics
Tiefeng Ma - Southwestern University of Finance and Economics
Shuangzhe Liu - University of Canberra
Kunio Shimizu - The Institute of Statistical Mathematics

Title On circular correlation for data on the torus

Abstract The topic of circular correlation for data on the torus is studied. Firstly, the order for two points on the circumference is considered and an order function is defined. Then, a moment coefficient to measure T-linear association between circular variables based on the order function is proposed. After the concordant on the torus is explained, a rank correlation coefficient is also proposed. A number of properties for the two coefficients are investigated and their comparisons with the existing alternatives are made. An example of wind direction data analysis is presented to illustrate our results.

Yoshinori Kawasaki - The Institute of Statistical Mathematics

Title Effective search for masked explanatory variables in linear regression

Abstract This paper presents an effective search for explanatory variables that correlate with response variable even being masked by other explanatory variables, which correspond to the explanatory variables that show no correlation but appear to be significant by including certain subset of explanatory variables in multiple regression model. Instead of all possible subset regression, the proposed search only requires pairwise Pearson's correlation coefficients between all explanatory variables and response variable, thereby enabling rapid computation. Multiple testing is utilized, followed by constructing shortest paths to response variable. A simple procedure for quantifying the potential correlations that are hidden is also presented. The proposed method is useful for explanatory analysis and works even under multicollinearity and when the number of explanatory variables is much larger than the sample size. Simulation studies demonstrate that the proposed method performs well for a high-dimensional model with severe multicollinearity, in which the elastic net fails. The practical usefulness is illustrated by a real environmental data. (This is a joint work with Dr. Masao Ueki of Kurume University.)

Keiichi Fukaya - The Institute of Statistical Mathematics

Title Accounting for misclassification errors in the estimation of transition probabilities: a spatial multistate dynamic site occupancy model

Abstract Inferring transition probabilities among ecological states is fundamental to community ecology because, based on Markovian dynamics models, these probabilities provide quantitative predictions about community composition and estimates of various properties that characterize community dynamics. Markov community models have been applied to sessile organisms because such models facilitate estimation of transition probabilities by tracking species occupancy at many fixed observation points over multiple periods of time. Estimation of transition probabilities of sessile communities seems easy in principle but may still be difficult in practice because resampling error (i.e., a failure to resample exactly the same location at fixed points) may cause significant estimation bias. Previous studies have developed novel analytical methods to correct for this estimation bias. However, they did not consider the local structure of community composition induced by the aggregated distribution of organisms that is typically observed in sessile assemblages and is very likely to affect observations.

In this study, a multistate dynamic site occupancy model is developed which can be used to estimate transition probabilities while accounting for resampling errors associated with local community structure. The model applies a nonparametric multivariate kernel smoothing methodology to the latent occupancy component to estimate the local state composition near each observation point, which is assumed to determine the probability distribution of data conditional on the occurrence of resampling error. By using computer simulations, it is confirmed that an observation process that depends on local community structure may bias inferences about transition probabilities. By applying the proposed model to a real dataset of intertidal sessile communities, it is also shown that estimates of transition probabilities and of the properties of community dynamics may differ considerably when spatial dependence is taken into account. Results suggest the importance of accounting for resampling error and local community structure for developing management plans that are based on Markovian models. The proposed model provides a solution to this problem that is applicable to broad sessile communities. It can even accommodate an anisotropic spatial correlation of species composition, and may also serve as a basis for inferring complex nonlinear ecological dynamics.

Kotaro Kagawa - The Institute of Statistical Mathematics

Title Under what conditions does hybridization promote adaptive radiation?

Abstract In adaptive radiation, many ecologically diverse species evolve in a single taxonomic group in relatively short term. For example, in the Lake Victoria in Africa, adaptive radiation of cichlid fish created about 500 endemic species within less than 100,000 years. Adaptive radiation typically occurs in the face of ecological opportunities that favor adaptation into various vacant ecological niches. Much evidence supports the view that adaptive radiation is driven by ecological speciation which is a mode of speciation where reproductive isolation between species develops through ecological differentiation. A longstanding question has been how genetic variation allowing rapid large-scale expansion of ecological traits has originated in adaptive radiations. Recently, growing body of empirical evidence suggests that homoploid hybridization may promote rapid ecological diversification in adaptive radiation. Hybridization can produce various novel genotypes within a few or tens of generations by mixing genomes from different parental lineages. Such hybrid genotypes may facilitate evolutionary invasion into novel ecological niches, which can lead to formation of new species. However, it has also been argued that hybridization can cause merge of parental species into a single hybrid swarm or lead to population collapse by disturbing adaptation. Thus, the role of hybridization in causing adaptive radiation is still unclear and seems to vary depending on situations.

The aim of this study is to reveal conditions in which hybridization promotes adaptive radiation. In this purpose, I developed an individual-based model to simulate adaptive radiation which is triggered by hybridization between two different lineages of a species. Individual-based model is a useful tool to simulate ecological and evolutionary dynamics of an ecosystem by tracking all individuals of organisms in the system. The model simulates the scenario that two parental lineages evolve independently in different sites for a certain period and then, they go into contact in another site where many ecological niches are available. Adaptive radiation takes place if evolution of many species utilizing various ecological niches occurs. To explore conditions under which hybridization promotes adaptive radiation, I conducted simulations with various conditions. Additionally, I simulated both scenarios that hybridization do and do not occur to test evolutionary roles of hybridization. Simulation results theoretically confirmed that hybridization can promote adaptive radiation and suggested that hybridization is particularly important in causing radiation when: (1) ecological niches are dissimilar one another, and (2) phenotypic effect size of mutations are small. Additionally, my simulation suggested that hybridization can promote radiation only when there is a moderate genetic distance between parental lineages. Contrarily, hybridization between highly genetically distant lineages tended to result in population collapse because excessive genotypic variation in hybrid populations disturbed adaptation. These results serve as the initial step toward theoretical understanding on evolutionary roles of homoploid hybridization.

ANU-UC-ISM Joint Symposium on Environmental Statistics 2016
University of Canberra, Australia
3 December 2016