Proceedings of ISM Symposium on Environmental Statistics 2016

【Date】January 15 (Friday), 2016
【Venue】The Institute of Statistical Mathematics (10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan), Auditorium (Second Floor)
【Organizers】K. Kanefuji(ISM), A. Welsh(ANU), K. Shimatani(ISM),
K. Shimizu(ISM)
【Host Organization】The Institute of Statistical Mathematics
【Joint Hosting】
Data Centric Science Research Commons, Research Organization of Information and Systems
NPO Japan Institute of Environmental Statistics

【Program】

Opening Address 10:10-10:20 **Tomoyuki Higuchi**, Director-General, ISM

Session 1 (Chair: **Yoshinori Kawasaki**, ISM) 10:20-11:40

10:20-11:00 Fitting misspecified linear mixed models
Alan Welsh, Mathematical Sciences Institute, ANU College of Physical & Mathematical Sciences, The Australian National University

11:00-11:40 A guided tour of the estimation and testing procedures involved in CRAD (Co-Regionalization Analysis with a Drift), with examples of applications in the environmental sciences
Pierre Dutilleul, Faculty of Agricultural and Environmental Sciences, Department of Plant Science, McGill University

Lunch 11:40-13:00

Session 2 (Chair: **Kenichi Shimatani**, ISM) 13:00-14:20

13:00-13:40 Sampling under constraints: some environmental applications
Louis-Paul Rivest, Département de Mathématiques et de Statistique, Faculté des Sciences et de Génie, Université Laval

13:40-14:20 Modelling the Relationship between Leaf Chemistry and Koala Population Density
Robert Clark, National Institute of Applied Statistical Research Australia, University of Wollongong

Break 14:20-14:35

Session 3 (Chair: **Shuangzhe Liu**, University of Canberra) 14:35-15:55

14:35-15:15 Hierarchical models do not need to be Bayesian
Brian Dennis, Department of Fish & Wildlife Sciences, University of Idaho

15:15-15:55 Weighted likelihood estimators for point processes and application in detecting spatial variations of seismicity clustering characteristics
Jiancang Zhuang, Department of Statistical Modeling, The Institute of Statistical Mathematics

Break 15:55-16:10

Session 4 (Chair: **Suhei Mano**, ISM) 16:10-17:30

16:10-16:50 A family of distributions for bivariate circular data
Shogo Kato, Department of Mathematical Analysis and
Statistical Inference, The Institute of Statistical Mathematics

16:50-17:30 Distributions on spheres defined by conformal transformation
Tomonari Sei, School of Information Science and Technology,
The University of Tokyo

Closing Address 17:30-17:35 **Alan Welsh**, The Australian National University

Fitting misspecified linear mixed models

Hwan-Jin Yoon [1] and **A. H. Welsh** [2]
The Australian National University, Canberra ACT 0200, Australia
[1] Statistical Consulting Unit,
and
[2] Mathematical Sciences Institute,
The Australian National University, Canberra ACT 0200, Australia

Abstract: Linear mixed models are widely used in a range of application areas, including ecology and environmental science. We study in detail the effects of fitting the two-level linear mixed model with a single explanatory variable that is misspecified because it incorrectly ignores contextual effects. In particular, we make explicit the effect of (the usually ignored) within-cluster correlation in the explanatory variable. This approach produces a number of unexpected findings. (i) Incorrectly omitting contextual effects affects estimators of both the regression and variance parameters not just, as is currently thought, estimators of the regression parameters and the effects are different for different estimators. (ii) Increasing the within cluster correlation of the explanatory variable introduces a second local maximum into the log-likelihood and REML criterion functions which eventually becomes the global maximum, producing a jump discontinuity (at different values) in the maximum likelihood and REML estimators of the parameters. (iii) Standard statistical software such as SAS, SPSS, STATA, lmer (from lme4 in R) and GenStat often returns local rather than global maximum likelihood and REML estimates in this very simple problem. (iv) Local maximum likelihood and REML estimators may fit the data better than their global counterparts but, in these situations, ordinary least squares may perform even better than the local estimators, albeit not as well as if we fit the correct model.

A guided tour of the estimation and testing procedures involved in CRAD (Co-Regionalization Analysis with a Drift), with examples of applications in the environmental sciences

Pierre Dutilleul
Faculty of Agricultural and Environmental Sciences,
Department of Plant Science, McGill University, Canada

Abstract: The abbreviation CRAD stands for “Co-Regionalization Analysis with a Drift”, a statistical method of spatial data analysis performed in two phases: Phase 1, to estimate drifts in a multivariate spatial dataset, and Phase 2, to analyze correlations from the residuals (i.e., after removal of the estimated drifts). Pelletier et al. (2009a, 2009b) introduced CRAD as an alternative to the original “Co-Regionalization Analysis” (CRA), in which the large-scale component of variability in the multivariate spatial dataset is considered random and modeled through a variogram (Goulard and Voltz, 1992; Goovaerts and Webster, 1994). Pelletier et al. (2009a, 2009b) did this after Larocque et al. (2007) had shown that the uncertainty associated with the estimation of parameters in the “Linear Model of Co-regionalization” (LMC) can be very high in CRA and Pelletier et al. (2004) had proposed a generalized least-squares estimation procedure for co-regionalization matrices in a LMC. On the testing side, Dutilleul and Pelletier (2011) defined tests of significance for structural coefficients of correlation in CRA and CRAD applications, and more recently, still in the spatial framework and following Dutilleul (1993, 2008) and Dutilleul et al. (2008), Dutilleul and Pelletier (unpublished) proposed a modified F-test of significance for the average R2 in “Redundancy Analysis” (RDA; Rao, 1964; van den Wollenberg, 1977; N.B.: The average R2 is the proportion of variance in the criterion variables that is reproducible linearly from the predictor variables in the RDA.). In this talk, we will navigate in the CRAD world. First, I will give a broad description of the framework and the main steps and their sequence. Then, I will explain in a good level of detail the procedures used for estimation and testing (e.g., by discussing underlying assumptions). Thereafter, the focus will be on the most recent testing results. And the talk will finish with the presentation and discussion of examples with environmental data.

A good part of the presented work is joint work with Bernard Pelletier (McGill

University).

References

Dutilleul, P., 1993. Modifying the t-test for assessing the correlation between two spatial processes. Biometrics 49, 305–314.

Dutilleul, P., 2008. A note on sufficient conditions for valid unmodified t testing in correlation analysis with autocorrelated and heteroscedastic sample data. Communications in Statistics – Theory and Methods 37, 137–145.

Dutilleul, P., Pelletier, B., 2011. Tests of significance for structural correlations in the linear model of coregionalization. Mathematical Geosciences 43, 819–846.

Dutilleul, P., Pelletier, B., A valid parametric test of significance for the average R2 in redundancy analysis with spatial data. Submitted for publication.

Dutilleul, P., Pelletier, B., Alpargu, G., 2008. Modified F-tests for assessing the multiple correlation between one spatial process and several others. Journal of Statistical Planning and Inference 138, 1402–1415.

Goovaerts, P., Webster, R., 1994. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland European. Journal of Soil Science 45, 79–95.

Goulard, M., Voltz, M., 1992. Linear coregionalization model: tools for estimation and choice of crossvariogram matrix. Mathematical Geology 24, 269–286.

Larocque, G., Dutilleul, P., Pelletier, B., Fyles, J.W., 2007. Characterization and quantification of uncertainty in coregionalization analysis. Mathematical Geology 39, 263–288.

Pelletier, B., Dutilleul, P., Larocque, G., Fyles, J.W., 2004. Fitting the linear model of coregionalization by generalized least squares. Mathematical Geology 36, 323–343.

Pelletier, B., Dutilleul, P., Larocque, G., Fyles, J.W., 2009a. Coregionalization analysis with a drift for multi-scale assessment of spatial relationships between ecological variables 1. Estimation of drift and random components. Environmental and Ecological Statistics 16, 439–466.

Pelletier, B., Dutilleul, P., Larocque, G., Fyles, J.W., 2009b. Coregionalization analysis with a drift for multi-scale assessment of spatial relationships between ecological variables 2. Estimation of correlations and coefficients of determination. Environmental and Ecological Statistics 16, 467–494.

Rao, C.R., 1964. The use and interpretation of principal component analysis in applied research. Sankhyā A: The Indian Journal of Statistics 26, 329–358.

van den Wollenberg, A.L., 1977. Redundancy analysis: An alternative to canonical correlation analysis. Psychometrika 42, 207–219.

Sampling under constraints: some environmental applications

Louis-Paul Rivest
Département de Mathématiques et de Statistique,
Faculté des Sciences et de Génie, Université Laval, Canada

Abstract: Environmental monitoring often involves the collection of data on a sample of units drawn randomly from a finite population. From a practical point of view, some samples might not be feasible; the selection of samples that meet operational constraints is discussed in this presentation. These constraints are implemented by balancing the sample on key explanatory variables. The so-called cube method of Deville and Tillé, and its generalizations, is used for that purpose. Two illustrations are presented. The first one is a creel survey for estimating the fishing effort for strip bass; a constraint is that data cannot be collected simultaneously at several sites as this is done by a single technician. The second one is a forest inventory where one would like to collect a sample of plots that are, at the same time, geographically clustered and representative of the forest diversity. This presentation is based on the paper "Incorporating spatial and operational constraints in the sampling designs for forest inventories" by Vallée, Ferland-Raymond, Rivest and Tillé, available as an early view on the Environmetrics website, see http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291099-095X/earlyview

Modelling the Relationship between Leaf Chemistry and Koala Population Density

Robert Clark
National Institute of Applied Statistical Research Australia,
University of Wollongong, Australia

Abstract: The koala's continued survival in many parts of Australia is uncertain. Such is the concern over the declines in koala populations in recent years that the Australian Government set up a Senate inquiry to investigate the issue and possible solutions. That inquiry ran for most of 2011 and released a report in September titled 'The koala—saving our national icon'. While the report proposes a number of actions to address the problem of declining koala numbers, a fundamental question remains regarding the distribution of koalas across eastern Australia. This presentation will describe the design and preliminary analysis of a study relating habitat characteristics, particularly leaf chemistry, to koala population density. The outcome variable, population density, was actually known prior to sampling areas, based on historical observation by ecologists. Explanatory variables, including tree species and chemical composition of leaves, were then collected by taking transect samples of trees, and intensively analysing leaf samples in the lab. Preliminary results are that tree subgenera is the most powerful predictor, probably mediated by the available nitrogen content of leaves. The analysis is complicated by the presence of spatial effects, zero densities, large extreme values, and loss of power due to aggregation to area level.

Hierarchical models do not need to be Bayesian

Brian Dennis
Department of Fish & Wildlife Sciences, University of Idaho, U.S.A.

Abstract: Hierarchical statistical models are not necessarily Bayesian, nor are they necessarily frequentist. Rather, statistical inferences for hierarchal models can be straightforwardly accomplished with Bayesian or frequentist approaches.

Hierarchical models are statistical models in which unobserved heterogeneity is modeled with probability distributions. Hierarchical models are being widely touted in the ecological and environmental sciences due to the tantalizing prospect of accounting for seemingly unlimited complexity in drawing conclusions from limited data. Random effects, missing data, unobserved covariates or latent variables, variable catch or sampling rates, and measurement errors in variables are examples of tough inference problems that have been tackled with hierarchical models.

In this presentation I will review recent developments in the uses of hierarchical models in ecology and provide a consumer's guide to statistical inferences. For many years, Bayesian approaches represented the only practical means of parameter estimation for hierarchical models, because the Markov chain Monte Carlo (MCMC) algorithms allowed posterior distributions to be simulated directly without having to calculate a likelihood function. However, it is now known that a simple tweak of the MCMC algorithms known as data cloning will redirect the Bayesian calculations toward providing maximum likelihood estimates of parameters along with other frequentist inferences (confidence intervals, hypothesis tests, AIC model selection, etc.). Therefore the choice of Bayesian or frequentist inference no longer needs to revolve simply around feasibility of calculation. Instead, the choice is returned to the investigator's philosophy of scientific method. That is as it should be, as Bayesian and frequentist approaches are wholly incompatible ways of drawing conclusions from data.

A dirty little secret of hierarchical models is parameter estimability (or rather, lack thereof). When building complex hierarchical models it is easy to pose models in which, for one reason or another, data do not inform well or provide much information about the values of parameters. The Bayesian framework, because it begins by assuming estimability of parameters, does not currently have satisfactory ways of diagnosing estimability problems.

However, the data cloning algorithm provides, along with ML estimates, a simple technique for diagnosing whether parameters (and functions of parameters) are estimable. The technique, a graph of eigenvalues, can be used by Bayesians and frequentists alike, but of course, the information so attained is in violation of Bayesian inferential principles (such as the likelihood principle).

I will provide a live demonstration of Bayesian and frequentist calculations for hierarchical models using the free WinBUGS software.

Weighted likelihood estimators for point processes and application in detecting spatial variations of seismicity clustering characteristics

Jiancang Zhuang
Department of Statistical Modeling,
The Institute of Statistical Mathematics, Tokyo, Japan

Abstract: Based on the technique of residual analysis, a weighted likelihood estimator for temporal and spatiotemporal point processes is proposed. Furthermore, we propose weighted Poisson likelihood estimators and weighted pseudo-likelihood estimators for spatial point processes. The weighted likelihood estimator is applied to the spatiotemporal Epidemic Type Aftershock Sequence (ETAS) model to study the spatial variations of seismicity characteristics in the Japan and the Italian regions.

A family of distributions for bivariate circular data

Shogo Kato
Department of Mathematical Analysis and Statistical Inference,
The Institute of Statistical Mathematics, Tokyo, Japan

Abstract: We propose a family of distributions for bivariate circular data. The proposed family is a bivariate extension of the distribution on the circle called the wrapped Cauchy distribution. The presented model is shown to have numerous appealing properties, amongst which figure: its five parameters have clear interpretations; its density is unimodal with a simple functional form; both its marginal and conditional distributions are wrapped Cauchy; random variate simulation from it is simple and efficient. It is seen that the proposed family is obtained by applying Möbius transformation to a pre-existing bivariate circular model. Method of moments and maximum likelihood estimation of its parameters are considered. An application of the proposed family to bivariate wind direction data is given.
This is joint work with Arthur Pewsey of the University of Extremadura, Spain; and M.C. Jones of the Open University, UK.

Distributions on spheres defined by conformal transformation

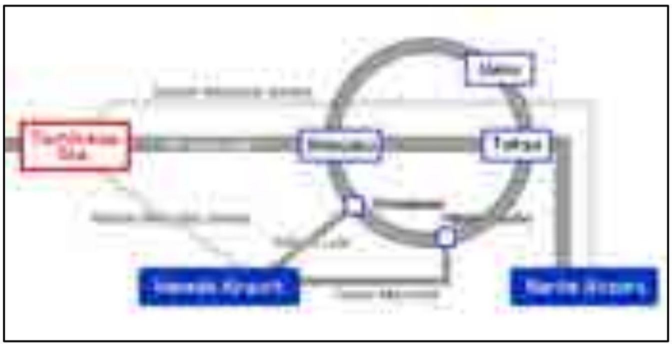
Tomonari Sei
School of Information Science and Technology,
The University of Tokyo, Tokyo, Japan

Abstract: It is known that for circular data the Moebius transform characterizes the wrapped Cauchy distribution (McCullagh 1996). In this talk, we generalize the argument to high-dimensional case. Some properties of density and likelihood functions are shown. This is joint work with K. Shimizu and K. Uesu.

Reference

McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation, The Annals of Statistics, 24, 787-808.

Address：
The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
Phone: +81-(0)50-5533-8500



Research Organization of Information and Systems
The Institute of Statistical Mathematics