Spatial Methods in Econometrics

Daniela Gumprecht

Dept of Statistics and Mathematics
Vienna University of Economics and Business Administration, Austria
e-mail: daniela.gumprecht@wu-wien.ac.at

In this talk I will give a brief overview of the characteristics of spatial data, why it is useful to use such data and how to use the information included in spatial data. The first question to be answered is: how to detect spatial dependency and spatial autocorrelation in data? Such effects can be found by calculating Moran's I, which is a measure for spatial autocorrelation, and using tests for spatial autocorrelation (Moran's test). Once we found some spatial structure we can use special models and estimation techniques. There are two famous spatial processes, the SAR- (spatial autoregressive) and the SMA- (spatial moving average) process, which are used to model spatial effects. For estimation there are mainly two different possibilities, the first one is called spatial filtering, where the spatial effect is filtered out and standard techniques are used, the second one is spatial two-stage least square estimation. Finally there are some results of a spatial analysis of R & D spillovers data (for 22 countries and 20 years) shown.

An Adaptive Hierarchical Test Procedure after Selecting Safe and Efficient Treatments

Franz König, Peter Bauer, Werner Brannath

Dept of Medical Statistics and Informatics
Medical University Vienna, Austria
e-mail: franz.koenig@meduniwien.ac.at

We consider the situation where during a multiple treatment (dose) control comparison high doses are truncated because of lack of safety and low doses are truncated because of lack of efficacy, e.g., by decisions of a data safety monitoring committee in multiple interim looks. We investigate the properties of a hierarchical test procedure for the efficacy outcome in the set of treatments carried on until the end of the trial, starting with the highest selected treatment to be compared with the control at the full level alpha. Left truncation, i.e., dropping doses in a sequence starting with the lowest dose, does not inflate the type I error rate. It is shown that right truncation does not inflate the type I error if efficacy and toxicity are positively related and treatment selection is based on monotone functions of the safety data. A positive relation is given e.g. in the case where the efficacy and toxicity data are normally disturbed with a positive pairwise correlation. Two specific right truncation rules are considered, one based on the mean treatment-control differences, the other based on the absolute treatment means at the interim looks. The power is increased if sample sizes saved for the truncated treatment groups are reallocated to the remaining treatments and control at the following stages.

Semi-Static Hedging Strategies for Exotic Options

Philipp Mayer

Dept of Mathematics
Graz University of Technology, Austria
e-mail: Philipp.Mayer@chello.at

The aim of hedging is to replicate the payoff of a non traded option. This allows to specify and to eliminate the inherent risk. When talking of hedging an exotic option mainly dynamic hedging strategies as the famous Black-Scholes delta-hedge are considered. However, recently static and semi-static hedging strategies became more popular, since they only involve a discrete number of trading times. Thus in the presence of transaction costs they might be superior to dynamic hedging. A short comparison of both approaches is drawn first. Classic semi-static hedging strategies for so called barrier options are presented next. Among those are the Derman-Egener-Kani algorithm and a method developed by Carr et al. for particular market models. Finally path-independent options are regarded. They can be replicated arbitrarily well by a portfolio of standard European options. This can also be used to hedge discretely monitored options as for example an Asian option.

Sampling Reconstruction of Stochastic Signals – The Roots in Fifties

Biserka Draščić

Dept of Sciences, Faculty of Maritime Studies
University of Rijeka, Croatia
e-mail: bdrascic@pfri.hr

Let us given a class of functions which are defined on some common domain. Can we find a discrete subset Λ of this domain such that every member of the class is determined uniquely by the collection of values that it takes on Λ and, if this is the case, how can we recover such a function completely using these "sampled values" only? This is the problem of sampling (analogue to digital transform) and reconstruction (digital to analogue transform).

A major application of sampling theory is in signal analysis. Here it provides the theoretical basis for modern pulse code modulation communication systems, having been introduced by Kotelńikov in 1933 and Shannon in 1949.

In this talk we are interested in the development of the sampling theory in signal analysis of *stochastic signals* in the fifties, especially in three most important papers for further development. First one is the famous paper by Balakrishnan from 1957 where he gave precise mathematical formulation of the Shannon's sampling principle with, what he initiated, the sampling reconstruction of band – limited weakly stationary stochastic processes.

The next two papers are actually the first two ones in which is considered sampling reconstruction of stochastic signals in the almost sure sense, or reconstruction "with probability 1". These are the papers by Belyaev and Lloyd, both from 1959. Here restoration in almost sure sense means that the Kotelńikov formula is valid for almost all sample functions of the considered stochastic signal.

In the talk one gives an overview of the mentioned articles and some further references on this subject for those who are interested will be listed.

An Estimation of Uniform Distribution if Data are Measured with Additive Error

Kristian Sabo

Dept of Mathematics
University of Osijek, Croatia
e-mail: ksabo@mathos.hr

We consider the problem of estimating the edges of the symmetric uniform distribution when data are measured with a normal additive error. The main purpose is to argument that the model is regular, as well as to give sufficient conditions for the existence of the maximum likelihood estimator and to suggest a numerical procedure for its computation. Also some generalizations of a similar problem in plane are studied. For this purpose we use some special numerical procedures for solving the nonlinear least squares problem.

Demometric Analysis of Croatian Population

Ana Štambuk

Economic Faculty
University of Rijeka, Croatia
e-mail: ana@efri.hr

In Croatia, same as in other European countries we experience decrease of fertility and demographic ageing of population. Level of fertility is below the replacement level in almost all countries. Decrease of fertility and increase of expected life expectancy leads to greater and greater share of elderly population and decrease of potential working contingent, which influence the entire economic development. Systematic analyzes of the population therefore is very essential. In order to analyze population as close as possible, we applied in our paper demometric methods. There is analyzed fertility, nuptiality, mortality and migration, as well as structure and population dynamics.

Besides usual demographic measurement we also calculated Bongaarts - Feeney quantum and tempo fertility and Coale - Trussel fertility method. Coale - McNeil nuptiality method was applied. The mortality is evaluated by Brass relational logit method. Migrations were evaluated by vitalostatic method. Besides economic and educational structures, stable - equivalent population structure was determined. As a potential for change of number of inhabitants also population momentum was calculated.

Demometric methods were also applied on other countries so that Croatian population was compared with population of different European countries.

On Speed of Stochastic Search on Decision Trees

Márton Ispány, Ilona Krasznahorkay

Dept of Applied Mathematics and Probability Theory
University of Debrecen, Hungary
e-mail: krasnil@inf.unideb.hu

The construction of decision trees is a commonly used and easily applied way of supervised learning. The aim is the prediction of a binary target variable on the basis of many predictor variables. This technique divides the field of predictors getting the target variable more and more homogeneous along the resulting partition. We modified the CART algorithm developed by Breiman et al. [1], improving with a stochastic search on the set of decision trees applying the Markov Chain Monte Carlo method. It was first proposed in a Bayesian framework by Chipman et al. [2].

We prove sharp rate of convergence for the Markov chain behind the algorithm. Namely, we show that $cn \log n$ steps are sufficient to reach the target distribution, where n is the number of the nodes in the corresponding decision tree. The technique of the proof is based on powerful estimation of the second largest eigenvalue developed in [3], [4] and [5].

References

[1] Breiman, L., Friedman, J. H., Olsen, A. O., Stone, C. J. (1984). *Classification and Regression Trees*, Wadsworth International Group.

[2] Chipman, H. A., George, E. I., McCulloch, R. E. (1998). *Bayesian CART Model Search*, Journal of the American Statistical Association, **93**, 443, 935-960.

[3] Jerrum, M. (1998). *Mathematical Foundations of the Markov Chain Monte Carlo Method.* Probabilistic Methods for Algorithmic Discrete Mathematics, Algorithm Combin., **16**, Springer, Berlin, 116-165.

[4] Jerrum, M., Sinclair, A. (1996). *The Markov Chain Monte Carlo Method: An Approach to Approximate Counting and Integration*, Approximation Algorithm for NP–hard Problems, (Dorit Hochbaum ed.), PWS.

[5] Diaconis, P., Holmes, S. P. (2002). *Random Walks on Trees and Matchings*, Electronic Journal of Probability, **7**, 6, 1-17.

Observable Operator Models

Ilona Spanczér

Dept of Mathematics
Technical University, Budapest, Hungary
e-mail: spanczer@math.bme.hu

If we investigate stochastic processes which have their own rules in the background then we have to choose a model. A new possibility is to use OOMs (Observable Operator Models). A widely used class of models for stochastic systems is hidden Markov models. In these models the Markov process is unknown but every state of the Markov process generate an observable output according to a probability distribution. We see only the outputs and we have to find out the model parameters based on the outputs. The OOMs are another view of this problem: instead of sequence of states we see sequence of operators according to the outputs. In the structure of OOM the initial distribution of the process is known and the operators generate the next output from the previous one. Using OOMs we would like to analyze stochastic processes and to predict the changes of the process.

Modelling Daily Water Discharges with Regime Switching Models

Krisztina Vasas

Dept of Probability Theory and Statistics
Eötvös Loránd University, Hungary
e-mail: v_krisz@cs.elte.hu

We give a talk on fitting a two-state autoregressive regime switching model for daily water discharge series registered at monitoring sites of River Tisza in Hungary. One peculiarity of the model is that the noise sequence switches distribution according to the gamma and normal law, the rising regime being governed by the gamma distributed part. When the change points of the regimes (which are hidden variables) are driven by a Markov chain, the estimation can be carried out by a simple implementation of the MCMC algorithm. In this case we update variables via Gibbs-sampling if it is possible, and Metropolis-Hastings algorithm otherwise. However, as regime lengths in hydrological series are known to deviate from the geometric distribution, Markov-modelling is not entirely satisfactory. When generalizations of the model are considered, more sophisticated estimation algorithms should be chosen. In some particular non-Markovian cases, we can still use the combination of the Gibbs-sampler and the Metropolis-Hastings method. This occurs when the length of the ascending period has negative binomial distribution, and the descending one is geometrically distributed. Another way of generalization if we let to alter the number of change points, then the estimation needs a so-called reversible jump MCMC method.

Keywords: Regime Switching; Reversible Jump MCMC; River Discharge Modelling

Inference from mitochondrial DNA data in forensic identification

Paola Berchialla

Dept of Public Health and Microbiology
University of Torino, Italy
e-mail: berkeley3@gmail.com

A general problem in forensic identification arises when a suspect is observed to have a genetic profile also known to be possessed by the offender whose mitochondrial DNA is recovered from a biological sample left at the scene of a crime. An immediate question is how much evidence against the suspect is provided by this matching.

The strength of the evidence against an incriminated individual is usually presented in the form of a likelihood ratio or its reciprocal (*profile match probability*).

The current method for estimating profile match probability is based on the frequency of sequences within databases. It would be the maximum likelihood estimate of the population proportion if observed sequences were treated as independent, ignoring the genealogical structure. On the other hand complete databases of reference populations have not been yet compiled and many sequences which occur in the population are not represented in the reference sample.

The aim of this paper is to develop a method for analyzing data that allows for the effect of the genealogical and mutational history which affect mitochondrial DNA molecule evolution and then computing the match probability in the framework of a fully likelihood based approach.

Conditional on the number of mutations, we consider the mutation process as a random walk between two alleles and assume that alleles can mutate at most once per generation. In this framework, which relies on the assumption that not all mutations occurring in the ancestry of a pair of genes lead to observable differences, a result is given for computing the profile match probability as a function of demographic and mutation parameters.

Finally, we introduce a hierarchical model for allele frequencies that allows for a mutation and a genealogical process based on the standard coalescence and a coalescent with growth model. This makes it possible to generate observations of mutation and demographic parameters from the post-data distribution in a simulation approach based on the acceptance-rejection method and calculate profile match probabilities via Monte Carlo method.

Scale-up Estimators in CATI Surveys for Estimating the Number of Choking Injuries in Children

Silvia Snidero[1], F. Zobec[2], Roberto Corradetti[1], Dario Gregori[3]

[1]Dept. of Statistics and Applied Mathematics, University of Torino, Italy
[2]S&A, Torino, Italy
[3]Dept. of Public Health and Microbiology, University of Torino, Italy
e-mail: snidero@econ.unito.it

The foreign body injuries in the upper aero-digestive ways are rare but not negligible events. Available data about foreign body injuries are those coming from the discharge records of the hospitals and from the death certificates. These data do not include the self-resolved injuries – those of minor severity – and indeed these cases are lost at observation. Thus, the overall injury rate is grossly underestimated.

It emerge the need of getting a reasonable estimate also of the non hospitalized cases. Common methods of probabilistic sampling are quite inadequate in this respect and better results are commonly obtained from non probabilistic sampling schemes.

Then, the idea is to estimate the number of all injuries with the scale-up method. This is a novel approach to estimate the size of hidden or hard to count subpopulations. Respondents are interviewed about the number of people known in several subpopulations (of known size) and a subpopulation E (which size is to be estimated).

Assuming that the proportion of subjects belonging to E over the number c of people in the social network of a person is the same that in the overall population we get the scale-up estimate of the size of the target subpopulation E.

We performed a CATI survey on a sample of about 1000 Italian women aged 18-50. 33 subpopulations of known size were chosen from Census and divided in two groups: populations with low sensitivity impact and high sensitivity impact.

Six different questionnaires were formed combining in different ways the target questions and the questions with low sensitivity and high sensitivity. The reason of these six different questionnaires was aimed at understanding how respondents react to each kind of question. Then, the results from the different questionnaires were compared.

Network Approach to Narrative Analysis

Vanja Govednik

University of Ljubljana, Slovenia
e-mail: vanja.ida@s5.net

Researchers have been studying narratives for several years trying to find some generalization of structure or existence of uniform rules in them. Vladimir Propp, when studying Russian fairy tales in 1928, discovered that there is a limited number of roles characters undergo and actions they perform. Benjamin Colby (1973) came to similar conclusions when dealing with Eskimo folktales. Several other researches in proceeding years dealt with narratives by help of different methods and from different scientific perspectives. The approach we were particularly interested in is the one parsing text of narratives into so called semantic triplets composed of subject, verb and object. We can read this kind of triplet as who did what to whom. This kind of parsing of text can be easily transformed into a network. Subjects and objects can be represented as vertices, and relations between them as arcs or edges. Because a narrative includes also time component, it determines a multi-relational temporal network. This kind of network can be further analyzed with the goal of finding some basic rules or patterns in it.

Goodness of Fit of Relative Survival Models

Maja Pohar

Dept of Medical Informatics
University of Ljubljana, Slovenia
e-mail: maja.pohar@mf.uni-lj.si

Relative survival methods are used in studies that aim to estimate cause-specific mortality, but do not have good information on the cause of death. The observed survival experience in a cohort of patients is compared to the expected or the population survival obtained from life tables. The commonly used regression models for relative survival are based on different basic assumptions and can give very different results. When evaluating the results of a model, we therefore need some goodness of fit information. In this presentation, we will focus on the additive and multiplicative regression model. While there is an abundance of methods to check the goodness of fit of the latter, no general methods exist for the additive model. We present a variety of procedures for testing goodness of fit that can be applied on either of the models. The methods are based on partial residuals defined analogously as the Schoenfeld residuals for the Cox model. With these residuals we can construct a process that converges to the Brownian motion and use its asymptotic distribution to check the goodness of fit.

Generalized Blockmodeling of Valued Networks

Aleš Žiberna

University of Ljubljana, Slovenia
e-mail: ales.ziberna@guest.arnes.si

The presentation several approaches to generalized blockmodeling of valued networks are presented. It is shown that the same criterion function used for generalized blockmodeling (Doreian, Batagelj, and Ferligoj, 2005) of binary networks can also be used for valued networks by specifying appropriate ideal blocks. Values on the ties are assumed to be measured on at least interval scale.

The first approach is a straightforward generalization of the generalized blockmodeling of binary networks proposed by Doreian, Batagelj, and Ferligoj (2005) to valued blockmodeling. The second approach is homogeneity blockmodeling. The basic idea of homogeneity blockmodeling is that the inconsistency of an empirical block to its ideal block can be measured by within block variability of appropriate values. What the appropriate values are is determined by the ideal block to which inconsistencies for selected empirical block are computed. These values are always based on values on the ties in the selected empirical block. New ideal blocks appropriate for blockmodeling of valued networks are presented together with the definitions of their block inconsistencies.

The advantages and disadvantages of proposed approaches are discussed. An example is also given.