Book of Abstracts

15th European Young Statisticians Meeting

September 10–14, 2007

Castro Urdiales, Spain

Local Organizing Committee

Inés Mª del Puerto, Chairman, *University of Extremadura*, Spain
Pedro Terán, Chairman, *University of Zaragoza*, Spain
Juan Romo, ERC of the Bernoulli Society, *Carlos III University of Madrid*, Spain
Miguel González, *University of Extremadura*, Spain
Rodrigo Martínez, *University of Extremadura*, Spain
Manuel Mota, *University of Extremadura*, Spain
Alfonso Ramos, *University of Extremadura*, Spain
Manuel Báez, *University of Extremadura*, Spain

International Organizing Committee

Philipp Pluch, Austria
Goele Massonnet, Belgium
Maroussia Bojkova, Bulgaria
Marek Omelka, Czech Republic
Zoran Vondracek, Croatia
Eva B. Vedel Jensen, Denmark
Cornelia Wichelhaus, Germany
Michael Zazanis, Greece
Krisztina Vasas, Hungary
Pierpaolo De Blasi, Italy
John Hinde, Ireland
Verica Bakeva, Macedonia
Pawel Zareba, The Netherlands
Lukasz Debowski, Poland
Antonia Turkman, Portugal
Marius Iosifescu, Romania
Andrey Selivanov, Russia
Barbora Arendacká, Slovakia
Tobias Ryden, Sweden
Sara van de Geer, Switzerland
Tetyana Kadankova, Ukraine
Jon Forster, United Kingdom

Contents

Welcome to 15th European Young Statisticians Meeting Castro Urdiales (Spain)

On behalf of the Local Organizing Committee I am pleased to welcome all of you to Castro Urdiales. The European Young Statisticians Meetings are held every two years. The idea of the meeting is to provide young researchers, perhaps just started in, or about to enter post-doctoral positions (and reasonably committed to remaining in research), an introduction to the international scene within the broad subject area. The venue of the European Young Statisticians Meeting is for the first time in Spain.

The contributions of the meeting keep a healthly balance between the theoretical and practical aspects of probability and statistics. All the talks are organized accordingly to 12 different topics in 17 sessions. There are no parallel sessions. In addition, there are two keynote lectures delivered by important scientists to which I want to thank their acceptation to participate in our Meeting.

On the other hand, the programme also offers cultural events that will allow us to discover the cities, gastronomy and sights of the dynamic and historical Spanish Communities of Cantabria and Basque Country as well as the charm and hospitality of their people.

I also would like to thank to all people that have actively participated in organizing the conference and to those entities and public administrations that have provided financial support. Particulary I want to express my acknowledgement to Professor Laureano González Vega, Director of the International Center of Mathematical meetings, for his inestimable help and worthy advice.

I wish you an enjoyable and successful meeting in Castro Urdiales, hoping that you will find the conference and the city a suitable setting to share your ideas and time.

Inés Mª del Puerto
Chair of the 15th EYSM

Schedule

	Monday, 10	Tuesday, 11	Wednesday, 12	Thursday, 13	Friday, 14
9:05-9:30	Registration	Presentation	Presentation	Presentation	Presentation
9:30-9:55	Opening	Presentation	Presentation	Presentation	Presentation
9:55-10:20	Keynote Lecture	Presentation	Presentation	Keynote Lecture	Presentation
10:20-10:45		Presentation	Presentation		Presentation
10:45-11:05	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Closing
11:05-11:30	Presentation	Presentation	Presentation	Presentation	
11:30-11:55	Presentation	Presentation	Presentation	Presentation	
11:55-12:20	Presentation	Presentation	Presentation	Presentation	
12:20-12:45	Presentation	Presentation	Presentation	Presentation	
12:45-15:25	Lunch	Lunch	Lunch[1]	Lunch	Lunch[2]
15:25-15:50	Presentation	Presentation	Excursion to Guggenheim museum in Bilbao	Presentation	
15:50-16:15	Presentation	Presentation		Presentation	
16:15-16:35	Coffee Break	Coffee Break		Coffee Break	
16:35-17:00	Presentation	Presentation		Presentation	
17:00-17:25	Presentation	Presentation		Presentation	
17:25-17:50	Presentation	Presentation		Presentation	
18:00-20:00	Guided Visit to Castro Urdiales				
20:00-	Welcome Drink			Closing Dinner	

[1] This day the lunch will be until 14:50, due to the excursion will begin at 15:00.

[2] It can be chosen to have lunch on Sunday, 9 or Friday, 14.

Monday, September 10	
	Keynote Lecture
9:55-10:45	**Daniel Peña** New Approaches for Analyzing Heterogeneity in Statistical Models
	Session 1: Hypothesis Testing I Chairman: Juan Romo
11:05-11:30	**Pierpaolo Brutti** On an Adaptive Goodness-of-Fit test with Finite Sample Validity for Random Design Regression Models
11:30-11:55	**Fragiadakis Kostas** Test of fit for symmetric variance gamma distributions
11:55-12:20	**Stefanie Hayoz** Behavior of nonparametric tests in longitudinal designs
12:20-12:45	**Gery Geenens** Nonparametric test for conditional independence in two-way contingency tables
	Session 2: Stochastic Models in Finance I Chairman: Petra Posedel
15:25-15:50	**Carl Lindberg** Robust portfolio optimization
15:50-16:15	**Ignacio Cascos** The financial risk of a set portfolio
	Session 3: Stochastic Processes I Chairman: Alfonso Ramos
16:35-17:00	**Amir Aliev** On the principle of smooth fit in optimal stopping problems
17:00-17:25	**Erika Fülöp** Statistical problems in a discrete time random field HJM type interest rate model
17:25-17:50	**Lars Andersen** Loss Rate Asymptotics

Tuesday, September 11	
	Session 4: Bayesian Statistics Chairman: Pierpaolo Brutti
9:05-9:30	**Roxana Ciumara** Convergence rate results for some empirical Bayes estimators
9:30-9:55	**Evgeny Burnaev** Quickest detection of intensity change for Poisson process in generalized Bayesian setting
9:55-10:20	**Matteo Ruggiero** Bayesian Nonparametric Construction of Fleming-Viot Models in Population Genetics
10:20-10:45	**Hannes Müller** Bayesian Transgaussian Kriging
	Session 5: Stochastic Models in Finance II Chairman: Ignacio Cascos
11:05-11:30	**Petra Posedel** Asymptotic analysis for a simple explicit estimator in Barndor-Nielsen and Shephard stochastic volatility models
11:30-11:55	**Ildikó Vitéz** Location as risk factor Spatial analysis of an insurance data-set
11:55-12:20	**Pavlo Shelyazhenko** Bounded arbitrage for multi-period model of financial market with discrete time
12:20-12:45	**Ivan Mitov** A discrete model for correlated default times and its application to CDO
	Session 6: Stochastic Processes II Chairman: Manuel Mota
15:25-15:50	**Iryna Rozora** Gaussian Process Simulation with application of the Theory of Square-Gaussian Processes
15:50-16:15	**Jaroslav Sevcik** Repairable systems with general repair
	Session 7: Resampling Chairman: Paulo Canas Rodrigues
16:35-17:00	**Claudia Kirch** Resampling methods for the change analysis of dependent data

Tuesday, September 11	
	Session 8: Density Estimation Chairman: Miguel González
17:00-17:25	**José E. Chacón** Bootstrap Bandwidth Selection Using an h-Dependent Pilot Bandwidth
17:25-17:50	**Julia Dony** Uniform consistency of kernel type estimators and conditional U statistics with general bandwidths

Wednesday, September 12	
	Session 9: Statistical Applications I Chairman: Andrew Parnell
9:05-9:30	**Patricia Geli** Modeling the Mechanism of Postantibiotic Effect
9:30-9:55	**Emma Holian** Mixture-Regression Cluster Model applied to Longitudinal Microarray Experiments
9:55-10:20	**Corinne Dahinden** Graphical Modeling for Discrete Random Variables with Application to Tissue Microarray (TMA) Experiments
10:20-10:45	**Robert Breitnecker** Analysing Regional Firm Startup Activity Using Geographically Weighted Regression: The case of Austria
	Session 10: Markov Processes I Chairman: Omiros Papaspiliopoulos
11:05-11:30	**Alina Nicolae** On asymptotic behavior of a finite Markov chain
11:30-11:55	**Sebastian Müller** Criteria for transience of branching Markov chains
11:55-12:20	**Krzysztof Latuszynski** A Regeneration Proof of the Central Limit Theorem for Uniformly Ergodic Markov Chain

Wednesday, September 12	
	Session 10: Markov Processes I Chairman: Omiros Papaspiliopoulos
12:20-12:45	**Vessela Stoimenova** On the statistics of branching processes with a random number of ancestors

Thursday, September 13	
	Session 11: Extreme Value Theory Chairman: Chen Zhou
9:05-9:30	**Marta Ferreira** Asymptotic and pre-asymptotic tail behaviour of a power max-autoregressive model
9:30-9:55	**Jan Dienstbier** Estimators of the extreme value index based on quantile regression
	Keynote Lecture
9:55-10:45	**Juan Antonio Cuesta Albertos** Random projections and goodness of fit tests for multidimensional data
	Session 12: Markov Processes II Chairman: Sebastian Müller
11:05-11:30	**Andrew Golightly** MCMC Sampling for Diffusion Processes
11:30-11:55	**Omiros Papaspiliopoulos** Recent advances in the simulation and inference for stochastic differential equation models
11:55-12:20	**Julie Lyng Forman** Estimating equations and inference from diffusion driven models
12:20-12:45	**Evangelos E. Vassiliou** Examples on lag distributed models subject to nonnegative divided differences of orders 2, 3 and 4

Thursday, September 13	
	Session 13: Hypothesis Testing II Chairman: José E. Chacón
15:25-15:50	**Jan Somorcik** Some multivariate multi-sample tests based on spatial medians
15:50-16:15	**Piotr Majerstki** Approximations to Most Powerful Invariant Tests for Multinormality
	Session 14: Random Matrices Chairman: Inés del Puerto
16:35-17:00	**Guangming Pan** Large dimensional random matrices
	Session 15: Time Series Chairman: Inés del Puerto
17:00-17:25	**Paulo Canas Rodrigues** Principal Component Analysis of Dependent Data
17:25-17:50	**Martina Hancova** Comparison of prediction quality of the best linear unbiased predictors in time series linear regression models

Friday, September 14	
	Session 16: Statistical Applications II Chairman: Rodrigo Martínez
9:05-9:30	**Andrew Parnell** Recent Advances in Palaeoclimate Reconstruction
9:30-9:55	**Chen Zhou** On spatial extremes: with application to a rainfall problem
	Session 17: Theory of Codes Chairman: Maroussia Bojkova
9:55-10:20	**Aleksandra Popovska** Random Codes
10:20-10:45	**Natasha Ilievska** Probabilistic models in quasigroup error-detecting codes

Keynote Lectures

New approaches for analyzing heterogeneity in statistical models

Daniel Peña, dpena@est-econ.uc3m.es
Department of Statistics. Universidad Carlos III de Madrid. Spain.

Abstract

New procedures for detecting heterogeneity in a sample with respect to a given class of models are presented. The relationship among the problems of outlier detection, cluster analysis, discriminant analysis and model selection is reviewed. It is shown that similar ideas can be used to solve this problems. Different approaches will be presented and compared in examples and Monte Carlo studies

Random projections and goodness of fit tests for multidimensional data

Juan Antonio Cuesta-Albertos, cuestaj@unican.es
Universidad de Cantabria, Spain.

Abstract

Counterexamples showing two different multidimensional distributions with a common marginal are very well known. The usual way to construct those counterexamples starts by fixing the marginal and, then, constructing two different distributions sharing this marginal. In [1] a different point of view is taken. There, the authors begin by having two multidimensional distributions P and Q, and, then, they consider the following problem: Given a continuous probability measure μ (for instance, gaussian), which is the μ-measure of the vectors, h, which satisfy that the (one-dimensional) marginals of P and Q along the line determined by h coincide? The answer is 1 if $P = Q$ and 0 if P and Q are different. Two sample goodness-of-fit tests follow straightforward from this result. Moreover, it was shown in [2] that this result can be extended to cover some families of distributions, thus providing ways to construct goodness-of-fit tests to those families. In particular, there it is shown that a distribution is gaussian if and only if almost every (one-dimensional) projection is gaussian. In this talk I will comment those results and will present some applications to real data sets.

References

[1] Cuesta-Albertos, J.A., R. Fraiman and T. Ransford (2007). *A sharp form of the Cramér-Wold theorem.* To appear in J. Theoret. Probab.

[2] Cuesta-Albertos, J.A., E. del Barrio, R. Fraiman and C. Matrán (2007). *The random projection method in goodness of fit for functional data.* To appear in Comput. Statist. and Data Anal.

Session 1

Hypothesis Testing I

On an Adaptive Goodness–of–Fit test with Finite Sample Validity for Random Design Regression Models

Pierpaolo Brutti, pierpaolo.brutti@uniroma1.it
"Sapienza" Università di Roma.

Keywords: Goodness–of–fit. Nonparametric Regression. Adaptive test. Random Design. Separation Rates. Warped Wavelets. U–statistics. Spherical needlets.
AMS: 62G08, 62G10

Abstract

Given an i.i.d. sample $\{(X_i, Y_i)\}_{i\in\{1,\ldots,n\}}$ from the random design regression model $Y = f(X) + \epsilon$ with $(X, Y) \in [0, 1] \times [-M, M]$, we consider the problem of testing the (simple) null hypothesis "$f = f_0$", against the alternative "$f \neq f_0$" from a non-asymptotic point of view and for a fixed $f_0 \in \mathsf{L}^2([0, 1], G_X)$, where $G_X(\cdot)$ denotes the (known) marginal distribution of the design variable X. The procedure proposed is an adaptation to the regression setting of a multiple testing technique introduced by Fromont and Laurent [2], and it amounts to consider a suitable collection of unbiased estimators of the L^2–distance $\mathsf{d}_2(f, f_0) = \int [f(x) - f_0(x)]^2 \mathrm{d}G_X(x)$, rejecting the null hypothesis when at least one of them is greater than its $(1 - u_\alpha)$ quantile, with u_α calibrated to obtain a level–α test. These estimators are built upon the *warped wavelet* system recently introduced by Picard and Kerkyacharian [1], and the resulting testing procedure turns out to be adaptive over a particular collection of approximation spaces linked to the classical Besov spaces. Possible extensions of the proposed procedure to other settings (e.g. two–sample problems, unknown design density, composite-hypotheses), and an analogous goodness–of–fit test on the sphere based on a new type of spherical wavelets, called needlets (see [3]) will also be sketched.

References

[1] Kerkyacharian, G. and Picard D. (2004). *Regression in random design and warped wavelets.* Bernoulli, 10(6), 1053–1105.

[2] Fromont M. and Laurent B. (2006) *Adaptive goodness–of–fit tests in a density model.* The Annals of Statistics, 34(2), 680–720.

[3] Narcowich, F., Petrushev P. and Ward J. (2006). *Decomposition of Besov and Triebel-Lizorkin spaces on the sphere.* J. Funct. Anal., 238, 530–564.

Test of fit for symmetric variance gamma distributions

Fragiadakis Kostas, kfragiad@econ.uoa.gr
Department of Economics, National and Kapodistrian University of Athens
8 Pesmazoglou Street, 105 59 Athens, Greece

Keywords: Variance Gamma Distribution, Characteristic Function, EM - type algorithm, Goodness of Fit

Abstract

The aim of this paper is to provide goodness-of-fit tests for the symmetric normal variance gamma distribution (SNVG). The NVG distribution is defined as a mixture of a normal distribution with a gamma distribution. The SNVG distribution is a three - parameter model, denoted by $SNVG(\delta, c, \lambda)$ which depends on a location parameter $\delta \in \mathbf{R}$, a scale parameter $c > 0$, and a shape parameter $\lambda > 0$.

Suppose that on the basis of independent copies $X_1, X_2, \ldots, X_n$, of a random variable X we wish to test the null hypothesis

H_0: The law of X is $SNVG(\delta, c, \lambda)$ for some $\delta \in \mathbf{R}$, $\sigma > 0$ and $\lambda > 0$.

We study a new family of omnibus tests of H_0 based on the empirical characteristic function (CF). Despite the fact that the density function of X is complicated, the CF, $\phi(t) = \mathbf{E}(e^{itX})$ of X is simply

$$\phi(t; \delta, c, \lambda) = e^{i\delta t}(1 + c^2 t^2)^{-\lambda}. \tag{1}$$

Its main characteristics are that the empirical CF is an unbiased and consistent estimator of the CF population and that, under specific sampling situations - including the present one - is more convenient to employ methods based on it, rather than classical methods such as moment based or likelihood inference methods. The "Fourier approach" has been proved to be a simple and very powerful tool for statistical inference, particularly in goodness - of - fit problems. Many inference procedures based on the empirical characteristic function have appeared lately in the literature.

A critical issue is the estimation of the parameters. Due to the complexity of the density function a direct maximization of the likelihood function is difficult while the method of moments estimates do not converge rapidly. In order to exceed this problem we use an EM - type algorithm.

The results of a Monte Carlo study would be presented. The Monte Carlo study was implemented by drawing 1000 samples of size $n = 50$, $n = 100$ and $n = 200$. However, since the null distribution of the test statistic depends on the value of the parameter λ, which is unknown, we resort to a parametric bootstrap procedure.

Behavior of nonparametric tests in longitudinal designs

Stefanie Hayoz, stefanie.hayoz@stat.unibe.ch
Institute of Mathematical Statistics and Actuarial Science.
University of Bern, Switzerland

Keywords: Nonparametric procedures. Longitudinal data. Repeated measurements. Simulation study.
AMS: 62G10, 62G35

Abstract

Nonparametric procedures need less model assumptions than the analogous parametric ones. Moreover the nonparametric procedures are more robust against deviations from the model assumptions. It is well known that the Friedman test is used to analyze the treatment effect in the randomized complete block design (see e.g. [3]). The Friedman test could also be used in a longitudinal design to analyze the time effect. Alternatives are an ANOVA type test (see e.g. [1]) and the parametric ANOVA test for repeated measures (see e.g. [2]). The three tests are compared with respect to the power and the control of the nominal level by means of a simulation study. This study incorporates different covariance structures are used to model the dependence among the repeated measurements and the results are based on 40'000 replications for each situation considering small sample sizes.

References

[1] Brunner, E. , Domhof, S. and Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments.* Wiley, New York.

[2] Davis, C. S. (1952). *Statistical methods for the analysis of repeated measurements.* Springer, New York.

[3] Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks.* Holden-Day, Inc., San Francisco.

Nonparametric test for conditional independence in two-way contingency tables

Gery Geenens, geenens@stat.ucl.ac.be
Institut de Statistique, Université catholique de Louvain, voie du Roman Pays, 20, 1348 Louvain-la-Neuve, Belgium

Keywords: contingency tables; conditional independence; chi-square test; nonparametric regression; long-range dependent process.
AMS: 62H17, 62E20, 62G08, 62G20, 62M07.

Abstract

Consider a two-way contingency table, built on two categorical variables R and S. A fundamental question in this context is whether R and S are independent or not. This over-studied testing problem mostly relies on the chi-square or the likelihood ratio test statistics. The principal drawback of this classical way of doing is that probabilities for an individual to fall in a cell of the table are assumed to be equal from one individual to another, so that we are not treating each individual as such but rather a group of supposed homogeneous units. This is often highly unrealistic, since in most of the practical situations some possibly known characteristics of each individual ought to influence or be associated with R, S, or both, and therefore influence the whole dependence structure of the table. In this case, a more judicious idea seems to analyze the conditional joint distribution of R and S given the vector of covariates, say X, and then to test for the conditional independence between R and S given X. In the literature, a striking fact is that the estimation of conditional probabilities associated with categorical responses, given a vector of covariates, is almost always treated via logistic regression methods, most of the time with very few validation of this parametric assumption. In this talk, we first present a nonparametric estimation procedure for the conditional probabilities to fall in each cell of the table. These estimates can be used as such, or be employed to validate a parametric assumption, like the logistic one. Secondly, we propose a generalization of the chi-square and the likelihood ratio tests to the case of conditional independence test, based on the above-mentioned nonparametric estimates of the conditional probabilities. The asymptotic law of the proposed test statistics is derived, and a simulation study illustrates the performance of the procedure in finite-sample situations.
Acknowledgements: Financial support from the "Interuniversity Attraction Pole", Phase VI from the Belgian Science Policy is gratefully acknowledged.

Session 2

Stochastic Models in Finance I

Robust portfolio optimization

Carl Lindberg, h.carl.n.lindberg@gmail.com
Weavering Capital, Sweden

Keywords: Black-Scholes model, robust portfolio optimization, Markowitz' problem, $1/n$ strategy, ranks.
AMS: 60H30,62P05

Abstract

It is widely recognized that when classical optimal strategies are used with parameters estimated from data, the resulting portfolio weights are remarkably volatile and unstable over time. The predominant explanation for this is the difficulty to estimate expected returns accurately. We propose to parameterize an n stock Black-Scholes model as an n factor Arbitrage Pricing Theory model where each factor has the same expected return. Hence the non-unique volatility matrix determines both the covariance matrix and the expected returns. This enables the investor to impose views on the future performance of the assets in the model. We derive an explicit strategy π^* which solves Markowitz' continuous time portfolio problem in our framework. The optimal strategy is to implicitly keep $1/n$ of the wealth invested in stocks in each of the n underlying factors. To illustrate the long-term performance of π^*, we apply it out-of-sample to a large data set. We find that it is stable over time and outperforms all the underlying market assets in terms of Sharpe ratios. Further, π^* had a significantly higher Sharpe ratio than the classical $1/n$ strategy.

The financial risk of a set portfolio

Ignacio Cascos, ignacio.cascos@uc3m.es
Dpto. de Estadística. Universidad Carlos III de Madrid.
Av. Universidad 30, E-28911 Leganés (Madrid). Spain.

Keywords: Coherent risk measure. Random set. Risk measure. Stochastic ordering.
AMS: 91B30, 91B82, 60D05

Abstract

We model a risky portfolio as a random set in the d-dimensional Euclidean space. A slight modification of the definition of risk measures for vector portfolios given by Cascos and Molchanov [4] makes is valid for the new set portfolios. In this framework, the risk of each set portfolio is given in terms of a subset of $\mathbb{R}^d$. Some relations between risk measures for set portfolios and stochastic orderings for random sets will be briefly discussed.
Acknowledgements: The financial support of the Spanish Ministry of Education and Science under grant MTM2005-02254 is acknowledged.

References

[1] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999) Coherent measures of risk. *Math. Finance*, 9:203–228.

[2] Bäuerle, N. and Müller, A. (2006) Stochastic orders and risk measures: consistency and bounds. *Insurance Math. Econom.*, 38:132–148

[3] Cascos, I. and Molchanov, I. (2003) A stochastic order for random vectors and random sets based on the Aumann expectation. *Statist. Probab. Lett.*, 63:295–305.

[4] Cascos, I. and Molchanov, I. (2007) Multivariate risks and depth-trimmed regions. *Finance Stoch.* In press.

[5] Jouini, E., Meddeb, M. and Touzi, N. (2004) Vector-valued coherent risk measures. *Finance Stoch.*, 8:531–552.

[6] Koshevoy, G. and Mosler, K. (1997) Zonoid trimming for multivariate distributions. *Ann. Statist.*, 25:1998–2017.

Session 3

Stochastic Processes I

On the principle of smooth fit in optimal stopping problems

Amir Aliev, amiraliev@gmail.com
Moscow State University, Faculty of Mechanics and Mathematics, Department of Probability Theory, 119992, Moscow, Russia.

Keywords: optimal stopping, smooth fit, free-boundary problems
AMS: 60G40

Abstract

Given a function G: $\mathbb{R}^n \to \mathbb{R}$ and a Markov process $(X_t)_{t\geq 0}$ consider the optimal stopping problem $V(x) = \sup_\tau \mathsf{E}_x G(X_\tau)$. It is well known that the optimal stopping time is given by $\tau^* = \inf\{t : X_t \in D\}$ where the set D is given by $D := \{x : V(x) = G(x)\}$ and V is the smallest superharmonic function that dominates G. This characterization could be reformulated as the free-boundary problem, where both V and D have to be determined. To get the explicit solution of the arising differential equations one has to impose additional boundary conditions on V. Often, this conditions follow from the principle of smooth fit that states $V' = G'$ on ∂D (in one-dimensional case) under various assumptions on G and X. This results are well-known and widely used, but only weakest of them could be extended directly to multi-dimensional case due to topological complications. Therefore, "smooth fit" is considered in topology weaker than standard topology of $\mathbb{R}^n$. In the article sufficient conditions are presented for this principle of smooth fit to hold in multi-dimensional case.
Acknowledgements: The author is grateful to professor Shiryaev for the formulation of the problem.

References

[1] Gikhman I. I. *and* Skorohod A. V. (1975). *The Theory of Stochastic Processes.* Nauka

[2] Grigelionis, B. I. *and* Shiryaev, A. N. (1966) *On Stefan's problem and optimal stopping rules for Markov processes.* Theory Probab. Appl. 11, 541-558

[3] Shiryaev, A. N. *and* Peskir, G. (2006). *Optimal Stopping and Free-Boundary Problems.* Birkhauser

Statistical problems in a discrete time random field HJM type interest rate model

Erika Fülöp and Gyula Pap, efulop@inf.unideb.hu
Department of Applied Mathematics and Probability Theory,
Faculty of Informatics, University of Debrecen,
Hungary

Keywords: HJM type forward interest rate model, strong consistency, LAN.

Abstract

We consider some statistical questions arising in a discrete time Heath-Jarrow-Morton (HJM) type forward interest rate model, where the interest rate curves are driven by a geometric spatial autoregression field. Such models were proposed by Gáll, Pap and Zuijlen [2].

Our aim is to test the autoregression parameter. We study strong consistency of the maximum likelihood estimator of the parameter ϱ. The difficulty is that the underlying sample consists of nonindependent random variables. Moreover, no explicit formula is available for the maximum likelihood estimators of ϱ.

In the stable ($|\varrho| < 1$) and unstable ($|\varrho| = 1$) cases we showed local asymptotic normality (LAN) of the sequence of the related statistical experiments in [1] in the sense of Le Cam [3], see also Van der Vaart [4]. The main gain of this result is that we obtain at once asymptotically optimal tests.

References

[1] Fülöp, E. and Pap, G. (2007). *Asymptotically optimal tests for a discrete time random field HJM type interest rate model.* Submitted to Acta Sci. Math.

[2] Gáll, J., Pap, G. and Zuijlen, M. v. (2006). *Forward interest rate curves in discrete time settings driven by random fields.* Comput. Math. Appl., 51(3-4), 387–396.

[3] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

[4] Vaart, A. W. van der (1998). *Asymptotic Statistics.* Cambridge University Press.

Loss Rate Asymptotics

Lars Nørvang Andersen, larsa@imf.au.dk
Department of Mathematical Sciences, University of Aarhus, Ny Munkegade,
DK-8000 Århus C, Denmark

Keywords: Lévy process, reflection, loss rate, asymptotics, heavy tail, asymptotics.
AMS: 60G51 62P30

Abstract

We consider a Lévy process $\{S_t\}$ which is reflected at 0 and $K > 0$. The reflected process $\{V_t^K\}$ is given as a solution to a Skorokhod problem, which implies a representation $V_t^K = V_0 + S_t + L_t^0 - L_t^K$, where $\{L_t^0\}$ and $\{L_t^K\}$ are the local times at 0 at K, respectively. The regenerative structure of $\{V_t^K\}$ yields a stationary distribution denoted π_K and the loss rate is defined as the mean of L_1^K in the stationary situation. The loss rate was studied in Asmussen & Pihlsgård [1], where it was expressed in terms of the characteristic triplet of $\{S_t\}$ and π_K, and asymptotics of the loss rate as $K \to \infty$ was derived in the case of negative drift and light tails. Asymptotics for positive drift is straightforward by reversing the role of the barriers 0 and K and using a conservation law.
We use the expression for the loss rate from [1] to derive asymptotics in the case of negative drift and heavy tails, as well as in the case of zero drift. In the zero drift case, functional limit theorems (with a Brownian or stable limit) play an important role and are based on continuity properties of the loss rate.

References

[1] Asmussen, S. and Pihlsgård, M. (2007) *Loss rates for Lévy processes with Two Reflecting Barriers* to appear in Mathematics of Operations Research.

[2] Asmussen, S. *Applied probability and queues* Springer, New York, 2003

Session 4

Bayesian Statistics

Convergence rate results for some empirical Bayes estimators

Roxana Ciumara, roxana_ ciumara@yahoo.com
Department of Mathematics. Calea Dorobantilor 15-17. Academy of Economic Studies, Bucharest. 010572 Romania.

Keywords: Empirical Bayes. Asymptotic optimality. Convergence rate.
AMS: 62C12

Abstract

The problem of asymptotic optimality and convergence rates for empirical Bayes estimators of some parameters was studied by different authors. Using a quadratic loss function, Liang [2] proved that, under certain conditions, the empirical Bayes estimator of the scale parameter in Pareto distribution is asymptotically optimal. Huang and Liang [1] analyzed the empirical Bayes estimator of the truncation parameter of truncated distribution family under asymmetric Linex loss. Later, Shi et al. [3] derived the convergence rate of empirical Bayes estimator for two-dimensional truncation parameters under Linex loss.

In this paper we evaluate the empirical Bayes estimators of the parameters of some truncated-type distributions, considering different classes of loss functions. We find the Bayes estimator and empirical Bayes estimator for the parameter taken into account. We establish conditions for asymptotic optimality and derive the associated rate of convergence to optimal risk. Finally, we consider a special case when the prior distribution is of a certain form.
Acknowledgements: The research is supported through CNCSIS Project 83/2007

References

[1] Huang, S.Y., Liang, T.C. (1997). *Empirical Bayes estimation of the truncation parameter with Linex loss.* Statistica Sinica, 7, 755-769.

[2] Liang, T.C. (1993). *Convergence rates for empirical Bayes estimation of the scale parameter in a Pareto distribution.* Computational Statistics and Data Analysis, 16, 35-45.

[3] Shi, Y., Gao, S., Shi, X. (2005). *Convergence rate of empirical Bayes estimation for two-dimensional truncation parameters under Linex loss.* Information Science, 171, 1-11.

Quickest detection of intensity change for Poisson process in generalized Bayesian setting

Evgeniy Burnaev, drazil@list.ru
Institute for Information Transmission Problems, Russian Academy of Sciences, Bolshoi Karetnyi per. 19, Moscow GSP-4, 127994, Russia.

Keywords: Poisson disorder problem. Quickest detection. Optimal stopping.
AMS: 60G40, 62M20

Abstract

We observe a stochastic process $X_t = \int_0^t I(s < \theta)dN_s^{\lambda_0} + \int_0^t I(s \geq \theta)dN_{s-\theta}^{\lambda_1}$, $X_0 = 0$, where $N_t^{\lambda_0}$ and $N_t^{\lambda_1}$ are Poisson processes with known intensities λ_0 and λ_1, $\lambda_0 > \lambda_1$, $I(\cdot)$ is an indicator function. Nonrandom unknown parameter $\theta \in [0, \infty]$ is a time when the "disorder" appears. Let $\tau = \tau(\omega)$ be a finite stopping time with respect to the filtration generated by the process X_t. We interpret τ as the decision that the disorder has happened at the time $\tau(\omega)$. Let P_t be the distribution of the process X_t under the assumption that the disorder happened at the time $\theta = t$. For every $T > 0$ we denote by $\mathcal{M}_T$ the set of stopping times with the mean time $T = \mathrm{E}_\infty\tau$ until the false alarm. For $\tau \in \mathcal{M}_T$ introduce the risk $B(T;\tau) = \frac{1}{T}\int_0^\infty \mathrm{E}_\theta(\tau-\theta)^+ d\theta$. The stopping time $\tau_T^* \in \mathcal{M}_T$ is called optimal if $B(T) = B(T;\tau_T^*) = \inf_{\tau\in\mathcal{M}_T} B(T;\tau)$. Let ψ_t be the Shiryaev's process, $\psi_t = \int_0^t \frac{L_t}{L_s}ds$, where $L_t = \exp\left(\log\left(\frac{\lambda_1}{\lambda_0}\right)X_t - (\lambda_1-\lambda_0)t\right)$ is the likelihood process. Denote $\rho = \frac{1}{2}\left(\frac{\lambda_1-\lambda_0}{\sqrt{\lambda_0}}\right)^2$, $\beta = \frac{1}{6}\left(\frac{\lambda_1-\lambda_0}{\sqrt{\lambda_0}}\right)^3$, $(-Ei(-x)) = \int_x^{+\infty}\frac{e^{-t}}{t}dt$ $(x > 0)$ – integral exponential function.

THEOREM. *For any $T > 0$ an optimal stopping time τ_T^* exists in $\mathcal{M}_T$ such that $\tau_T^* = \inf\{t \geq 0 : \psi_t \geq T\}$. Moreover, the generalized Bayesian risk $B(T) = \frac{V(0;T)}{T}$, where $V(x;T)$ solves differential-difference equation $(1-(\lambda_1-\lambda_0)x)V'(x;T) + \lambda_0(V(x\frac{\lambda_1}{\lambda_0};T) - V(x;T)) = -x$, $V(T;T) = 0$. If $\lambda_0 \to \infty$ and $\lambda_1 \to \infty$ such that $\lambda_1 = \lambda_0 + C\sqrt{\lambda_0}$ for some fixed constant $C < 0$, then $B(T) = B_0(T) + \varepsilon B_1(T) + \ldots$, where $\varepsilon = 1/\sqrt{\lambda_0}$, $B_0(T) = \frac{1}{\rho}\left[e^{\frac{1}{\rho T}}\left(-Ei\left(-\frac{1}{\rho T}\right)\right) - \left(1 - \frac{1}{\rho T}\int_0^\infty \frac{e^{-t}\log(1+\rho T\cdot t)}{t}dt\right)\right]$, $B_1(T) = \frac{\beta}{\rho^2}\left[\frac{3}{\rho T}\int_0^\infty \frac{e^{-t}\log(1+\rho T\cdot t)}{t}dt + e^{\frac{1}{\rho T}}Ei\left(-\frac{1}{\rho T}\right)\cdot\left(\frac{1}{2\rho T}-1\right) - \frac{5}{2}\right]$.*

The author is grateful to Professor A.N. Shiryaev for his help.

Acknowledgements: The research was partially supported by the analytical departmental target programme RNP.2.2.1.1.2467, the Program of State Support of Leading Scientific Schools HS-5379.2006.1, the Fundamental Research Program of the RAS Presidium 15, grant RGNF 06-02-91821 a/G, grant RFFI 05-01-00944.

Bayesian Nonparametric Construction of Fleming-Viot Models in Population Genetics

Matteo Ruggiero, matteo.ruggiero@unito.it
Department of Statistics and Applied Mathematics. Piazza Arbarello, 8. University of Turin. 10122 Torino, Italy.

Keywords: Fleming-Viot process; Measure-valued diffusion; Generalised Pòlya-urn schemes; Stationary distribution; Two-parameter Poisson-Dirichlet process
AMS: Primary: 60G57; 60J35. Secondary: 60J60; 92D25.

Abstract

The Fleming-Viot process is a probability-measure-valued diffusion which arises as the large population limit of a wide class of population genetics models. In a few of its formulations its stationary distribution is known to be the Dirichlet process or a mixture of Dirichlet processes, but its connections with Bayesian statistics are still to be explored.

This work provides several explicit constructions of Fleming-Viot processes in the Bayesian nonparametric framework, and yields a previously unknown stationary distribution. In particular, by means of known and newly defined generalised Pòlya-urn schemes, several types of pure jump particle processes are introduced, describing the evolution in time of an exchangeable population. In each case, the process of empirical measures of the individuals converges in distribution in the Skorohod space to a specific Fleming-Viot diffusion, and the stationary distribution is shown to be the de Finetti measure of the infinite sequence of individuals. In presence of viability selection the stationary distribution turns out to be the two-parameter Poisson-Dirichlet process.

Acknowledgements: Matteo Ruggiero is supported by CRT Foundation, Turin, and by MIUR, grant 2006/133449.

Bayesian Transgaussian Kriging

Hannes Müller, hannes.mueller@uni-klu.ac.at
Institut für Statistik. Alpen Adria Universität Klagenfurt, Austria

Keywords: Geostatistics; Kriging; Bayesian Statistics

Abstract

In geostatistics a widely used method for prediction is kriging. It is well known and used for many years. But some limits are inherent on the traditional ways of kriging, i.e. simple, ordinary and universal kriging.

Kriging is based on the assumptions that the covariance function is exactly known and the underlying random field is a gaussian field. In practice, neither the trend or the variogram are exactly known, but on the other hand there may be expert knowledge for the trend or the variogram that should be used for prediction in geostatistics. So a mixture of kriging and Bayesian statistics can be useful.

In theory, Baysian transgaussian kriging can handle random fields with non-gaussian behaviour and with various trends and different link functions, but the problem is the computational effort. Many applications need near real-time evaluations of the random process, so the CPU-time is limited.

In the article and in the presentation a short overview on the wide field of kriging and Bayesian statistics is given, showing some of the current research the author is doing in moment.

Session 5

Stochastic Models in Finance II

Asymptotic analysis for a simple explicit estimator in Barndorff-Nielsen and Shephard stochastic volatility models

Petra Posedel, pposedel@efzg.hr
Department of Mathematics, Faculty of Economics and Business Zagreb, University of Zagreb, Croatia.

Keywords: Martingale estimating functions, stochastic volatility models with jumps, consistency and asymptotic normality

Abstract

We provide a simple explicit estimator for discretely observed Barndorff-Nielsen and Shephard models, prove rigorously consistency and asymptotic normality based on the single assumption that all moments of the stationary distribution of the variance process are finite, and give explicit expressions for the asymptotic covariance matrix.

We develop in detail the martingale estimating function approach for a bivariate model, that is not a diffusion, but admits jumps. We do not use ergodicity arguments.

We assume that both, logarithmic returns and instantaneous variance are observed on a discrete grid of fixed width, and the observation horizon tends to infinity. This anaysis is a starting point and benchmark for further developments concerning optimal martingale estimating functions, and for theoretical and empirical investigations, that replace the (actually unobserved) variance process with a substitute, such as number or volume of trades or implied variance from option data. This is joint work with Friedrich Hubalek, Technische Univesität Vienna.

References

[1] Barndorff-Nielsen, Ole E. and Shephard, N. (2001). *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics.* J. R. Stat. Soc. Ser. B Stat. Methodol., 63(2), 167-241.

[2] Sørensen, M. (1999). *On asymptotics of estimating functions.* Brazilian J. Prob. and Stat., 13, 111-136.

[3] Jun Pan, (2002). *The jump-risk premia implicit in options: evidence from an integrated time-series study.* J. Financial Economics, 63, 3-50.

Location as risk factor. Spatial analysis of an insurance data-set

Ildikó Vitéz, vildiko@cs.elte.hu
Department of Probability Theory and Statistics. Eotvos Lorand University. Hungary.

Keywords: Spatial statistics, Claim fequency, Markov random field, Potts model, Monte Carlo EM.

Abstract

Our aim was to examine the territorial dependence of risk for household insurances. We assumed that the counts of claims of each contracter are Poisson distributed with the appropriate parameter. Besides the classical risk factors such as type of wall, building, etc., we considered the location associated to each contract. In our first approach we used Markov random field model to describe the spatial effect. Basically there are two ways of fitting the model; we fit a GLM to the counts of claims with the classical risk factors and regarding their effects as fixed we fit the spatial model. Alternatively we can estimate the effects of all covariates jointly. Although this latter may seem to be a more accurate approach, its high complexity and computational demands makes it no feasible in our case. To overcome the disadvantages of the separate estimation of the classical and the spatial risk factors we proceeded as follows: first we used a GLM for the non-spatial covariates, and then fitted the spatial model by MCMC. Then we refitted the GLM and the spatial model again, and we iterated this procedure several times. We achieved a much better fit by performing these two steps eight times. Another approach was that instead of using hierarchical Bayes model we estimated the value of the spatial parameters by a maximum likelihood estimate. In this model we assumed Potts model for the relative risk of the regions. As result we achieved analytically insolvable equations, so we used Expectation Maximization (EM) algorithm to get the estimates. In case of high dimension we also used Monte Carlo EM algorithm.

References

[Dimakos and Frigessi di Rattalma (2002)] Dimakos, X.K., Frigessi di Rattalma, A., 2002: Bayesian premium rating with latent structure, *Scandinavian Actuarial Journal*, (2002), 162-184.

[Green, P. J., Richardson S. (2002)] Green, P. J., Richardson S. , Hidden Markov Models and Disease Mapping, American Statistical Association

Bounded arbitrage for multi-period model of financial market with discrete time

Yuliya Mishura, Georgiy Shevchenko and Pavlo Shelyazhenko,
Pavlo.Shelyazhenko@gmail.com
Department of Mechanics and Mathematics, Kyiv Taras Shevchenko National University,
64 Volodymyrska, 01033 Kyiv Ukraine

Keywords: Arbitrage. ε-arbitrage. Financial market. Multi-period market. Equivalent measure. Fundamental theorem of asset pricing.

Abstract

In a general discrete time market model with portfolio constraints, an ε-arbitrage opportunity is an opportunity to make a riskless profit of amount at least ε while trading a limited amount of assets. In a one-period framework with contingent initial data and in a multi-period framework we prove no-ε-arbitrage criteria similar to the classical fundamental theorem of asset pricing. Classical result is that no-arbitrage is equivalent to the existance of an equivalent martingale measure, i.e. discounted price process is a martingale w.r.t. this measure. We prove that there is no ε-arbitrage iff there exists such an equivalent measure with a bounded density that the discounted price process is in certain sense close to a martingale w.r.t. this measure.

References

[1] Dalang R.C., Morton A., Willinger W. (1990). *Equivalent martingale measures and no-arbitrage in stochastic securities market models.* Stochastics and Stochastics Reports, 29, 185-201.

[2] Mishura Yu.S. (2003). *The fundamental theorem of financial mathematics for bounded arbitrage* (Ukrainian). Applied Stat., Actuarial and Financial Math., 1-2, 49-54

A discrete model for correlated default times and its application to CDO

Ivan Mitov, imitov@yahoo.com
Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski",
Sofia, Bulgaria

Keywords: Default times. Default correlation. Dependence. Portfolio Loss Distribution. CDO.
AMS: 62P05

Abstract

A simple discrete time model for the dependence between the default times of assets (firms) in a given portfolio is discussed. The model is based on multivariate geometric distributions. A possible way for calibration of the model is discussed too. In the framework of this model portfolio loss distribution is calculated and a Monte Carlo based estimation of the expected losses in a given CDO is done.
Acknowledgements: This work is partially supported by NFSI grant No.VU-MI-105/2005.

References

[1] K. Giesecke, (2002), *An exponential model for dependent defaults*, (working paper), Humboldt - Universitat zu Berlin.

[2] D. Duffie and N. Garleanu, (2001), *Risk and Valuation of Collateralized Debt Obligations.*, Financial Analyst's Journal, 57, 41-59.

[3] D. Li, 2000, *On default correlation: A copula approach*, Journal of Fixed Income 9(4), 43-54.

[4] P. Schonbucher and D. Schubert, 2001, *Copula-dependent default risk in intensity models*, Working paper, Bonn University.

[5] D. Wong, 2000, *Copula from the limit of multivariate binary model*, (working paper) Quantitative Finance, Bank of America Corporation.

[6] Fan Yu, 2003, *Dependent default in intensity-based models* , (working paper), University of California, Irvine.

Session 6

Stochastic Processes II

Gaussian Process Simulation with application of the Theory of Square-Gaussian Processes

Iryna Rozora, irozora@bigmir.net
Department of Applied Statistics, Faculty of Cybernetics, Kyiv National Taras Shevchenko University, Ukraine

Keywords: Square-Gaussian processes, simulation, model construction, Gaussian processes, reliability, accuracy
AMS: 60G60, 68U20, 65C20, 60G15

Abstract

In many applied areas which use theory of stochastic processes the problem arises to estimate probability that a random vector process $\overrightarrow{X}^T(t) = (X_1(t), X_2(t), \ldots, X_d(t))$ leaves some region on some interval of time. For example, some systems break off on the interval $[0, T]$ if

$$\sup_{0 \le t \le T} \sum_{k=1}^{d} c_k^2(t) X_k^2(t) > \varepsilon$$

or, in more general case, if

$$\sup_{t \in T} \overrightarrow{X}^T(t) A(t) \overrightarrow{X}(t) > \varepsilon,$$

where ε is a sufficiently large number, (T, ρ) is a metric space, $\overrightarrow{X} = (\overrightarrow{X}(t), t \in T)$ is a process that generates the system, $A(t)$ is a matrix (in most cases positive semidefinite). The process $\overrightarrow{X}(t)$ may be considered as Gaussian due to the central limit theorem. Thus the problem arises to estimate the probability

$$P\left\{\sup_{t \in T} \overrightarrow{X}^T(t) A(t) \overrightarrow{X}(t) > \varepsilon\right\},$$

or the probability
$$P\left\{\sup_{t \in T} |\overrightarrow{X}^T(t) A(t) \overrightarrow{X}(t) - \mathbf{E}\overrightarrow{X}^T(t) A(t) \overrightarrow{X}(t)| > \varepsilon\right\},$$
where $\overrightarrow{X}(t)$ is a Gaussian vector process and $A(t)$ is a symmetric matrix. The process $\overrightarrow{X}(t)$ is considered as centered one.

The estimates of such large deviation probability for square-Gaussian stochastic processes are obtained. This result is used for model construction of Gaussian stochastic process as input process of same system, taking into account output process, with given reliability and accuracy in Banach space.

Repairable systems with general repair

Jaroslav Sevèík, sevcik@karlin.mff.cuni.cz
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic. Pod Vodárenskou vìží 4, Praha 8, 182 08, Czech Republic.

Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University Prague. Sokolovská 83, Praha 8, 186 75, Czech Republic.

Keywords: Repairable system. Virtual age. Hazard Rate. Maintenance policy.
AMS: 60K05, 90B25

Abstract

The contribution deals with general repair models for repairable system. Various ways of modeling the impact of repairs on a system condition are examined. The most common is to assume the repairs impact the failure intensity following a virtual age process (of the general form) proposed by Kijima. Another option considers repairs performing the time-dependent scale transformation of the governing distribution function. Ultimately, repairs directly reducing system deterioration represented by cumulative hazard rate are taking into account. In the second part of the article preventive maintenance actions of repairable system are discussed.
Acknowledgements: The research was supported by the GAÈR grant 201/05/H007.

References

[1] Kijima, M. (1989). *Some results for repairable systems with general repair.* Journal of Applied Probability, 26, 89-102.

[2] Finkelstein, Max S. (2000) *Modeling a process of non-ideal repair.* Recent advances in reliability theory (Bordeaux, 2000), Stat. Ind. Technol., 41-53.

[3] Kahle, Waltraud and Love, Charles E. (2003) *Modeling the influence of maintenance actions.* Mathematical and statistical methods in reliability (Trondheim, 2002), Ser. Qual. Reliab. Eng. Stat., 7, 387-399.

Session 7

Resampling

Resampling methods for the change analysis of dependent data

Claudia Kirch, ckirch@mathematik.uni-kl.de
Technical University Kaiserslautern. Department of Mathematics.
Erwin-Schrödinger Str. 67663 Kaiserslautern.

Keywords: mean change, block resampling, dependent observations
AMS: 62G10,62G09

Abstract

The fundamental question in change-point analysis is whether an observed stochastic process follows one model or whether the underlying model changes at least once during the observational period. This talk focuses on detecting an abrupt mean change in a sequence of dependent data.

In change-point analysis critical values for testing procedures are usually obtained by distributional asymptotics. These critical values, however, do not sufficiently reflect dependency. Moreover it is a well-known fact that convergence rates especially for extreme-value statistics are very slow. Using resampling methods we obtain better approximations, which take possible dependency structures more efficiently into account.

We focus on a block permutation method which splits the observations $X(1), \ldots, X(n)$ into L blocks of length K. Instead of resampling $X(\cdot)$, we resample the blocks $X(Kl+1), \ldots, X(K(l+1)), l = 0, \ldots, L-1$, but keep the order within the blocks. Then we prove that the original statistics and their resampling counterparts follow the same distributional asymptotics, thus we can use the quantiles of the bootstrap distribution as critical values.

Some simulation studies illustrate that the permutation tests usually behave better than the original tests if performance is measured by the $\alpha-$ and $\beta-$errors, respectively.

References

[1] Antoch, J., and Hušková, M. (2001). *Permutation tests for change point analysis.* Statist. Probab. Lett., 53, 37-46.

[2] Kirch, C. (2007). *Block permutation principles for the change analysis of dependent data.* J. Statist. Plann. Inference, 137, 2453-2474.

Session 8

Density Estimation

Bootstrap Bandwidth Selection Using an h-Dependent Pilot Bandwidth

José E. Chacón, jechacon@unex.es
Departamento de Matemáticas. Avda. de Elvas, s/n. Universidad de Extremadura. 06071 Spain.

Keywords: bootstrap bandwidth choice, h-dependent pilot bandwidth, kernel density estimation, plug-in method
AMS: 62G07, 62G20

Abstract

The problem of choosing the bandwidth h for kernel density estimation is considered (see [2]). All the plug-in and bootstrap bandwidth selection methods require the use of a pilot bandwidth g and the usual way to make an h-dependent choice of g is by obtaining their asymptotic expressions separately and solving the two equations for the sample size n. In contrast, we obtain the asymptotically optimal value of g for every *fixed* h, thus making our selection "less asymptotic". Exact error expressions show that some usually assumed hypotheses have to be discarded in the asymptotic study in this case. Based on this idea, two versions of a new bootstrap-type bandwidth selector are proposed, and their properties are analyzed through theoretical results and a simulation study. The new selector can achieve the best possible rate with the minimal asymptotic variance, and the simulations show that in practice one version of this selector provides a good compromise between cross-validation and the popular Sheather-Jones method (see [1]): it is not so variable as cross-validation and its centerpoint is significantly less biased than that of the Sheather-Jones bandwidth selector in most cases.
Acknowledgements: This research has been supported by the Spanish Ministerio de Ciencia y Tecnología project MTM2006-06172.

References

[1] Sheather, S. J. & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **53**, 683–690.

[2] Wand, M. P. & Jones, M. C. (1995). *Kernel smoothing.* Chapman & Hall, London.

Uniform consistency of kernel–type estimators and conditional U–statistics with general bandwidths

Julia Dony, jdony@vub.ac.be
Department of Mathematics, Vrije Universiteit Brussel,
Pleinlaan 2, B–1050 Brussels, Belgium.

Keywords: empirical processes, kernel estimation, Nadaraya–Watson, nonparametric estimation, regression, uniform in bandwidth consistency, U–statistics.
AMS: 60F15, 62G08

Abstract

Let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random variables and denote by f_X the (unknown) marginal density function of X. For practical use, it is often important to be able to estimate this density, as well as the regression function $m(t) = \mathbb{E}[Y|X = t]$. A popular class of estimators consists of kernel–type estimators, which have been intensively studied for a long time. Some well–known examples are the Nadaraya–Watson estimator and the local polynomial regression function estimators.

We start by giving some recent uniform consistency results for kernel–type estimators of f_X and $m(t)$, with a concise overview of the used methodology. The most recent results are shown to be valid uniformly in bandwidth as well, implying that the consistency is preserved when the bandwidth is chosen depending upon the datatset and/or the location. This kind of result is both from theoretical as practical point of view a considerable improvement.

Next, we look at conditional U–statistics, which were introduced by Stute to estimate the general regression function $m_\varphi(\mathbf{t}) = \mathbb{E}[\varphi(Y_1, \ldots, Y_m)|(X_1, \ldots, X_m) = \mathbf{t}]$, and for which pointwise consistency was proved in 1991. We are now interested in the uniform convergence of these estimators, so a strong uniform consistency result is established, where a uniformity in the bandwidth is considered as well. In addition, the corresponding convergence rates can be derived from this result.

References

[1] Dony, J.,Einmahl, U. and Mason, D.M. (2006). *Uniform in bandwidth consistency of local polynomial regression function estimators.* Austr. J. Stat. **35,** 105–120.

[2] Dony, J. and Mason, D.M. (2007). *Uniform in bandwidth consistency of conditional U–statistics.* Submitted.

[3] Stute, W. (1991). *Conditional U–statistics.* Ann. Probab. **19**, 812–825.

Session 9

Statistical Applications I

Modeling the Mechanism of Postantibiotic Effect

Patricia Geli, patricia@math.su.se
Mathematical Statistics. Department of Mathematics. Stockholm University.
SE-10691 Stockholm. Sweden

Keywords: Antibiotic resistance. Kolmogorov equations. Penicillin Binding Proteins. Postantibiotic effect
AMS: 60J80

Abstract

Models for the emergence and spread of resistance have many challenges. In this paper we look at the specific problem of modeling the postantibiotic effect (PAE), the delayed regrowth of the bacteria after complete removal of an antibiotic.

A stochastic model for describing one of the possible underlying biological mechanisms of PAE is formulated. The model is based on the theory of penicillin binding proteins (PBPs), where the PAE is the time required by the bacteria to synthesize new PBPs before growth. Newly synthesized PBPs are unsaturated and becomes saturated under antibiotic pressure and eventually removed due to death.

The model assumes that unsaturated PBPs are attached (synthesized) to a bacterium according to a Poisson process and that these in turn are saturated with an intensity proportional to the antibiotic concentration of the treatment. The calculations and results are divided into three simplifying steps toward a more realistic approach. At first, we assume constant antibiotic concentration and no initial PBPs. Secondly, we assume constant antibiotic concentration, but with an initial set of unsaturated PBPs (no saturated PBPs). Thirdly, we assume exponentially declining antibiotic concentration and the same initial set of unsaturated PBPs.

The stochastic models are solved using a set of Kolmogorov equations and exact solutions with interesting properties can be derived for all three steps. The results are useful for giving a better understanding of the time properties of PAE.

Mixture-Regression Cluster Model applied to Longitudinal Microarray Experiments

Emma Holian, emma.holian@ul.ie
Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland.

Keywords: Microarray; Longitudinal; Mixtures; Regression; Random effects.
AMS: 60J80

Abstract

The aim of this work is to explore various statistical techniques to identify genes which contribute to some change in phenotype level. For example, the response of fish kept under stressful conditions for various lengths of time. We aim to assess the level of *differential* expression of each gene in the tissue samples and also attempt to model the expression patterns of genes over time, not only to classify genes by similarities in expression patterns, but also to model these patterns as specified functions.

The proposed Mixture-Regression Cluster Model is developed to model *and* cluster the genes into groups according to their expressions measured over time. This model is similar to that of the multivariate normal mixture model in that clusters are identified by the EM algorithm but is adapted to incorporate the flexibility of regression curves to fit the trends. In this way, additional features such as covariates, random effects and correlation structures can be incorporated into the model while potentially offering a considerable saving on the number of parameters required to model the trends.
Acknowledgements: Special Thanks to colleagues working at the National Diagnostics Centre, and in the Mathematics Department of National University of Ireland, Galway, where the work was carried out.

References

[1] Wit, E., and McClure, J. (2004). *Statistics for Microarrays: design, analysis and inference.* John Wiley & Sons.

[2] Diggle, P.J., Heagerty, P., Liang, K., and Zeger, S.L. (2002). *Analysis of Longitudinal Data.* Oxford University Press.

[3] McLachlan, G.J., and Peel, D. (2000). *Finite Mixture Models.* New York: John Wiley & Sons.

Graphical Modeling for Discrete Random Variables with Application to Tissue Microarray (TMA) Experiments

Corinne Dahinden, dahinden@stat.math.ethz.ch
Seminar für Statistik, ETH Zürich and Competence Center for Systems Physiology and Metabolic Diseases, University/ETH Zürich

Keywords: Graphical models, Hierarchical models, Interactions, Lasso, Log-linear models, Variable selection
AMS: 68R01, 62P10

Abstract

Tissue microarrays (TMA) are composed of hundreds of tissue sections from different patients arrayed on a single glass slide. With the use of immunohistochemical staining, they provide a high-throughput method of analyzing potential biomarkers on large patient samples. The assessment of the expression level of a biomarker is usually performed by the pathologist on a categorical scale.

The analysis of the interaction of these biomarkers and in particular the estimation of the graph structure associated with the underlying discrete random variables, are of biological importance. Questions such as how the interaction pattern changes with progressing tumor grade or with survival time are of direct biological interest. However, the estimation of the interaction structure requires sophisticated techniques. Our approach is to fit an ℓ_1-regularized log-linear model assuming a multinomial sampling scheme in order to obtain the graphical model. The regularization becomes necessary as after cross-tabulation of the samples in contingency tables, many cell entries remain zero, leading to so-called sparse contingency tables, where standard procedures fail to work.

We compare our approach with other methods for graphical modeling. Moreover, biological validation of the estimated interaction structure is done by mapping to known biochemical pathways.

Analysing Regional Firm Startup Activity Using Geographically Weighted Regression: The case of Austria

Robert Breitenecker, robert.breitenecker@uni-klu.ac.at
Klagenfurt University
Department of Innovation Management and Entrepreneurship
Universitätsstrasse 65-67, 9020 Klagenfurt, Austria.

Keywords: spatial models, geographically weighted regression, founding activity
AMS: 91B72

Abstract

One subject in entrepreneurship research is to find factors which explain regional variations in firm startup activity. Reviewing the entrepreneurship literature points out that in most studies classical linear models are used to identify influencing factors on regional founding rates and that the spatial component of data is mostly disregarded. Estimating a global model assumes that the influence of the independent variables are constant over the whole study region. I will test in this paper if the global linear model is appropriate to estimate the founding activity. Therefor global and local indicators for spatial autocorrelation (Moran I, [1]) are calculated first, to show the spatial dependence in regional firm startup rates and additional regional data on the level of 93 counties of Austria. Further a geographically weighted regression (GWR) model is estimated to test whether a constant relationship between the founding rate and dependent variables exists over the whole study region [2]. The test of spatial dependence results in significant positive autocorrelation for all regional variables. The local parameter values of the GWR model indicate significant regional variations which indicates that a global linear model is not appropriate to explain the startup activity in Austria.

References

[1] Anselin, L. (1995). *Local Indicators of Spatial Association-LISA*. Geographical Analysis, 27(2), 93–115.

[2] Fotheringham, S. A., Brunsdon, C., und Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, Chichester.

Session 10

Markov Processes I

On asymptotic behavior of a finite Markov chain

Alina Nicolae, alinanicolae@unitbv.ro
Department of Mathematical Analysis and Probability. University Transilvania of Braşov. Romania.

Keywords: difference equation; finite Markov chain; infinite product; simulated annealing; uniform weak ergodicity; uniform strong ergodicity; convergence.
AMS: 60J10

Abstract

We consider a finite nonhomogeneous Markov chain with transition matrices $(P_n)_{n\geq 1}$ such that $\lim_{n\to\infty} P_n = P$, where P has $p \geq 1$ irreducible and aperiodic closed classes and, perhaps, transient states. We give sufficient conditions for uniform weak and uniform strong ergodicity and convergence, respectively, of the nonhomogeneous chain in terms of similar properties of a nonhomogeneous Markov chain of smaller size. We show the dependence of long-run behavior of $(P_n)_{n\geq 1}$ on spectral properties of the matrix P and on the nature of the perturbations $V_n = P_n - P$, $\forall n \geq 1$. Our conditions are compared to some other results existing in literature. Markov chains of this type occur in simulated annealing, a stochastic algorithm for global optimization.

References

[1] Karlin, S. and Taylor, H.M. (1975). *A First Course in Stochastic Processes.* Academic Press. New York.

[2] Isaacson, D.L. and Madsen, R.W. (1976). *Markov Chains:Theory and Applications.* Wiley. New York.

[3] Iosifescu, M. (1980). *Finite Markov Processes and Their Applications.* Wiley, Chichester and Editura Tehnică, Bucureşti.

[4] Nicolae, A. (2007). *A sufficient condition for the convergence of a finite Markov chain.* Math. Reports. 59 (3), 2007.

Criteria for transience of branching Markov chains

Sebastian Müller, mullerse@math.uni-muenster.de
Institut für mathematische Strukturtheorie, TU Graz

Keywords: branching Markov chains, transience, recurrence, superharmonic functions
AMS: 60J10, 60J80

Abstract

A branching Markov chain (BMC) is a system of particles in discrete time. The BMC starts with one particle in an arbitrary starting position x. At each time particles split up in offspring particles independently according to some probability distributions that may depend on the locations of the particles. The new particles then move independently according to a Markov Chain.

An irreducible Markov Chain is either recurrent or transient: either all or none states are visited infinitely often. It turns out that this dichotomy breaks down for BMC and that one can classify BMCs in three different types. Let $\alpha(x)$ be the probability that, starting the BMC in x, the state x is hit infinitely often by some particles. There are three possible regimes: *transient* ($\alpha(x) = 0\ \forall x$), *weakly recurrent* ($0 < \alpha(x) < 1\ \forall x$) and *strongly recurrent* ($\alpha(x) = 1\ \forall x$). We give equivalent criteria for transience of BMC and discuss some interesting consequences.

A Regeneration Proof of the Central Limit Theorem for Uniformly Ergodic Markov Chains

Krzysztof Latuszynski, latuch@gmail.com
DDepartment of Mathematical Statistics. Institute of Econometrics. Warsaw School of Economics. 02-554 Warsaw. Poland.

Keywords: Markov Chain, CLT, Uniform Ergodicity, Regeneration.
AMS: 60J05

Abstract

Central limit theorems for functionals of general state space Markov chains are of crucial importance in sensible implementation of Markov chain Monte Carlo algorithms as well as of vital theoretical interest. Different approaches to proving this type of results under diverse assumptions led to a large variety of CTL versions. However due to the recent development of the regeneration theory of Markov chains, many classical CLTs can be reproved using this intuitive probabilistic approach, avoiding technicalities of original proofs. In this paper we provide a regeneration proof of a CLT for functionals of uniformly ergodic Markov chains, thus solve the open problem posed in [3]. Moreover we discuss the difference between one-step and multiple-step small set condition.
The presentation is joint work with Witold Bednorz from Warsaw University.

References

[1] Cogburn, R., 1972, *The Central Limit Theorem for Markov Processes.* In Le Cam, L. E., Neyman, J. & Scott, E. L. (Eds) Proc. Sixth Ann. Berkley Symp. Math. Sttist. and Prob., 2, 458-512.

[2] Meyn S. P., Tweedie R. L., 1993. *Markov Chains and Stochastic Stability.* Springer-Verlag.

[3] Roberts G. O., Rosenthal J. S., 2005. *General state space Markov chains and MCMC algorithms.* Probability Surveys 1:20-71.

On the statistics of branching processes with a random number of ancestors

Vessela Stoimenova, stoimenova@fmi.uni-sofia.bg
Sofia University,
Faculty of Mathematics and Informatics,
Department of Probability, Operations Research and Statistics,
5 J. Boucher Str. 1164 Sofia, Bulgaria

Keywords: offspring mean, individual distribution, robust estimation
AMS: 60J80

Abstract

In the present paper we consider the robust estimation in the Bienayme - Galton - Watson processes with a random number of ancestors (BGWR processes) in the sense of the weighted and trimmed likelihood. The classical estimators of the individual mean and distribution are compared with their robust modifications via simulations and computational results.
Acknowledgements: This research was partially supported by the National Science Fund of Bulgaria, Grant No VU-MI-105/2005.

Session 11

Extreme Value Theory

Asymptotic and pre-asymptotic tail behaviour of a power max-autoregressive model

Marta Ferreira, msferreira@math.uminho.pt
Department of Mathematics of School of Sciences . Campus de Gualtar. University of Minho. 4710-057 Braga Portugal.

Keywords: Markov Chains. Extreme value theory. Stationary sequence. Dependence conditions. Tail index. Extremal index. Tail dependence index. Asymptotic independence.
AMS: 60G70

Abstract

Nowadays we can not ignore the demand of various areas like hydrology, geophysics or finances, on modeling extreme data or exceedance data above certain high levels. The assumption of independent observations gave place to a more realistic statement: the temporal dependence. Phenomenon therein are often associated with markovian sequences for which we claim simple ways of modeling in what concerns the extremes topic, like for instance, *max-autoregressive* processes (Alpuim [1] and Canto e Castro [3]). Here we present a max-autoregressive process involving a power transformation - $ARMAX_p$ - whose parameter $c \in (0,1)$ relates directly with the coefficient of tail dependence (η) of Ledford and Tawn [4]. This index characterizes the penultimate tail dependence of a process and can be related with a threshold-dependent extremal index, which assumes an important role when extending discussions of extreme values from independent and identically distributed (i.i.d.) sequences to stationary ones (Bortot and Tawn, [2]). In this paper, we analyze some asymptotic extremal features, we study the local dependence structure and we state the threshold-dependent extremal index connection with parameter η, for the $ARMAX_p$ process.

References

[1] Alpuim, M.T. (1989). *Contribuições à teoria de extremos em sucessões dependentes*. Ph. D. Thesis, FCUL.

[2] Bortot, P. and Tawn, J. (1998). Models for the extremes of Markov Chains. *Biometrika* **85**, 851-867.

[3] Canto e Castro, L. (1992). *Sobre a Teoria Assintótica de Extremos*. Ph. D. Thesis, FCUL.

[4] Ledford, A. and Tawn, J.A. (1997). Modelling Dependence within joint tail regions. *J. Royal Statist. Soc.Ser. B* **59** 475-499.

Estimators of the extreme value index based on quantile regression

Jan Dienstbier, dienstbi@karlin.mff.cuni.cz
Department of Probability and Statistics, MFF UK, Solovská 83, 186 75, Praha 8, Czech Republic

Keywords: extreme value index, quantile regression
AMS: 60G70

Abstract

In the contribution we discuss an estimation of the extreme value index in the case of linear regression model. We propose estimators based on the extreme regression quantiles slightly expanding older results of Chernozhukov [1]. Our aim is to describe a general class of estimators of the extreme value index based on regression quantiles. We show consistency and asymptotic normality of estimators belonging to this class using methodological tools based on Dress (see [2], [3]) and Gutenbrunner [4]. Various conditions and their necessity are discussed. We demonstrate a performance of proposed estimators on a small simulation study.
Acknowledgements: This work was supported by Czech Republic Grant GAČR 201/05/H007.

References

[1] Chernozhukov V. (2005) Extremal Quantile Regression, Annals of Statistics, Vol. 33, 2, 806–839

[2] Drees, H. (1998a) On smooth statistical tail functionals, Scand. J. Statist., 25, 187–210.

[3] Drees, H. (1998b) A general class of estimators of the extreme value index, J. of Stat. Planning and Inference, 55, 95–112

[4] Gutenbrunner, C., Jurečková J., Koenker R. and Portnoy, S., (1993) Tests of linear hypotheses based on regression rank scores, Nonparametric Statistics, Vol. 2, 307–331

Session 12

Markov Processes II

MCMC Sampling for Diffusion Processes

Andrew Golightly, a.golightly@ncl.ac.uk
School of Mathematics & Statistics, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK.

Keywords: Stochastic Differential Equation, Markov chain Monte Carlo, Particle Filter

Abstract

It is well known that likelihood inference for arbitrary nonlinear diffusion processes observed at discrete times is problematic since closed form transition densities are rarely available. One widely used solution adopts the treatment of Pedersen [1] and involves the introduction of latent data points between every pair of observations to allow an Euler-Maruyama approximation of the true transition densities to become accurate. In recent literature, Markov chain Monte Carlo (MCMC) methods have been used to sample the posterior distribution of latent data and model parameters; however, naive schemes suffer from a mixing problem, highlighted by Roberts and Stramer [2], that worsens with the degree of augmentation. We will consider some recently developed MCMC schemes that are not adversely affected by the amount of augmentation. The methodology will be illustrated by estimating parameters governing the diffusion approximations of some interesting systems biological models.
Acknowledgements: This research is joint work with Darren J. Wilkinson, School of Mathematics & Statistics, Newcastle University, UK

References

[1] Pedersen, A. (1995). *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations.* Scandinavian Journal of Statistics 22, 55-71.

[2] Roberts, G.O. & Stramer, O. (2001). *On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm.* Biometrika 88, 603-621.

Recent advances in the simulation and inference for stochastic differential equation models

Omiros Papaspiliopoulos, o.papaspiliopoulos@warwick.ac.uk
Department of Statistics, University of Warwick, Coventry, CV4 7AL UK.

Keywords: Exact simulation, Monte Carlo maximum likelihood, filtering, EM algorithm

Abstract

In this talk I will review recent advances in the exact simulation of diffusion processes, introduced mainly in [1], [2] and [3]. I will present a general methodology which can be used to solve a variety of statistical problems related to diffusion processes and I will give special attention to an EM algorithm for the estimation of parameters and unobserved states.

References

[1] Beskos, A., Papaspiliopoulos, O. and Roberts, G.O. (2006). *Exact Simulation of Diffusion Sample Paths with Applications.* Bernoulli, 12, 1077-1098.

[2] Beskos, A., Papaspiliopoulos, O., Roberts, G.O. and Fearnhead, P. (2006). *Exact and computationally efficient likelihood–based inference for discretely observed diffusions.* J. R. Statist. Soc. B., 68, 1-25.

[3] Fearnhead, P., Papaspiliopoulos, O. and Roberts, G.O. (2006). *Particle filters for partially observed diffusions.* Submitted.

Estimating equations and inference from diffusion driven models

Julie Lyng Forman, jufo@biostat.ku.dk
Department of Biostatistics. University of Copenhagen. Denmark.

Keywords: general estimating function, goodness of fit testing, stochastic differential equation.

Abstract

Diffusion-type models provide a natural and flexible framework for modeling phenomena that evolve continuously and randomly with time. However their statistical analysis is complicated. Only rarely the functional form of the likelihood function of a diffusion is explicitly known. The same problem of course occurs with diffusion driven models such as summed diffusions, integrated diffusions, and stochastic volatility models. General estimating equations often present a simple alternative means for fitting these models. I will illustrate the use of estimating equations for diffusion-type models driven by diffusions having linear drift and quadratic squared diffusion coefficient. I shall use the term Pearson diffusions as the invariant distributions of this class of diffusions belong to the Pearson system. Futher I will present a new goodness of fit test based on estimating equations for varying sampling frequencies. The basic idea is to compare the parameter estimates based on the full data to those obtained from downsamples. The test can be used to distinguish different kinds of diffusion driven models.
Acknowledgements: The research presented here is joint work with Michael Sørensen, Department of Mathematical Sciences, University of Copenhagen and with Bo Markusen and Helle Sørensen, Department of Natural Sciences, LIFE, University of Copenhagen.

Examples on lag distributed models subject to nonnegative divided differences of orders 2, 3 and 4

Evangelos E. Vassiliou, evagvasil@econ.uoa.gr
Department of Economics, University of Athens

Keywords: approximation, divided differences, lag distributed models, least squares, quadratic programming, r-convexity.
AMS: 60J80

Abstract

We consider noisy measurements from a time series that follow a linearly distributed lag model. It is usual to assume that the lag coefficients lie on some curve and then specify the curve by a least squares calculation. However, we define the r-th order smoothness priors by requiring nonnegative divided differences of order r for the lag coefficients. Such priors do not imply any parameterization of the lag curve and provide a more accurate representation of the prior knowledge. For the calculation of the solution we propose an algorithm that gives the least squares change to the data subject to nonnegative divided differences of the lag coefficients of order r, where r is a prescribed positive integer. The problem is a strictly convex quadratic programming calculation, where each of the constraints functions depends on r+1 adjacent components of the smoothed values of the lag coefficients. We take account of this special structure and use a special active set method that is more efficient than general quadratic programming algorithms. In fact we construct a basis that reduces the equality-constrained minimization calculations that appear during the quadratic programming iterations to unconstrained minimization ones, which depend on much fewer variables. Finally, we present an example that illustrates our approach.
Acknowledgements: The work of the author is supported by research grand 70/7/7145 (Heracleitus) of the university of Athens, which is co-financed within Op. Education by the European Social Fund and National Resources. We feel grateful to them all.

References

[1] Almon, S. (1965). *The distributed lag between capital appropriations and expenditures.* Econometrica, Vol. 33, No. 1., 178-196.

[2] Demetriou I. C., Lipitakis E. A. & Vassiliou E. E. (2006). *Efficient use of Toeplitz Matrices for Least Squares Data Fitting by Nonnegative Differences.* Hermis-$\mu\pi$, Volume 6, 135-153.

[3] Shiller, R. J. (1973). *A distributed lag estimator derived from smothness priors.* Econometrica, Vol. 41, No. 4., 775-788.

Session 13

Hypothesis Testing II

Some multivariate multi-sample tests based on spatial medians

Jan Somorcik, somorcik@fmph.uniba.sk
Department of Applied Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, 842 48 Bratislava, Slovakia.

Keywords: Spatial median. Spherical and elliptical symmetry. Asymptotic equivalence.
AMS: 62H11

Abstract

The multivariate multi-sample location problem in case of spherically or elliptically symmetric distributions is considered and we test the hypothesis about equality of the location parameters (see e.g. [1]). We present some analogies of the well-known Lawley-Hotelling test statistic based on spatial medians of the samples.

In case of spherical symmetry, asymptotic distribution of the median-based test statistics is derived under the hypothesis and also under the sequence of Pitman's alternatives. Using the Bahadur-type representation of the spatial median (see [2]) the asymptotic equality of the median-based statistics to an obvious multi-sample extension of the well-known Rayleigh test statistic is proved under the hypothesis.

Some Monte Carlo results are introduced, involving also the case when the underlying distribution is elliptically symmetric.
Acknowledgements: The research was supported by the VEGA grant 1/3016/06 of the Scientific Grant Agency of the Slovak Republic.

References

[1] Um, Y. and Randles, R. H. (1998). *Nonparametric tests for the multivariate multi-sample location problem.* Statistica Sinica, 8, 801-812.

[2] Chaudhuri, P. (1992). *Multivariate location estimation using extension of R-estimates through U-statistics type approach.* Annals of Statistics, 20, 897-916.

Approximations to Most Powerful Invariant Tests for Multinormality

Piotr Majerski, majerski@agh.edu.pl
Faculty of Applied Mathematics. AGH University of Science and Technology,
30-059 Kraków, Poland.

Keywords: Most powerful invariant test. Laplace approximation.
AMS: 62H15

Abstract

We consider the problem of testing multinormality against alternatives invariant with respect to some subgroup of affine transformations. In [3], a general form of the most powerful invariant (MPI) test has been obtained. Unfortunately, applicability of the MPI test is rather limited, due to complicated, intractable integrals. With the aid of the Laplace method for integrals, we derive large sample approximations for the MPI tests. The cases of bivariate exponential and uniform alternatives are studied in details, whereas higher dimensional extensions are outlined. It is shown in the both bivariate cases, that a further approximation for the Laplace approximation can be given. This leads to the likelihood ratio (LR) test statistic. A final conclusion is that the likelihood ratio test statistic can be seen as a formal expansion of the MPI test statistic, with a known upper bound for the relative error of the approximation. The Monte Carlo simulation study shows, that powers of both, the Laplace approximation, as well as the LR test are very close to the power of the most powerful invariant test even in small sample sizes.
Acknowledgements: The research was partly supported by the KBN local grant No 11.420.04

References

[1] Ducharme, G.R., Frichot, B. (2003), *Quasi-most powerful invariant goodness-of-fit tests*, Scand. J. Statist., 30, 399-414;

[2] Majerski, P. (2007), *Approximations to Most Powerful Invariant Tests for Multinormality*, To be published;

[3] Szkutnik Z. (1988). *Most powerful invariant tests for binormality.* Ann. Statist., 16, 292-301.

Session 14

Random Matrices

Large dimensional random matrices

Guangming Pan, stapgm@gmail.com
Eurandom, PO Box 513 - 5600 MB Eindhoven, The Netherlands.

Keywords: Sample covariance matrix. eigenvalues and eigenvectors. wireless communication
AMS: 15A52;60F15

Abstract

This paper focuses on the theory of spectral analysis of Large Random Matrix. Concerning eigenvalues and eigenvectors some important results and methods are reviewed and moreover, some latest results regarding eigenvectors are also presented. Finally, the application of random matrix to wireless communication is illustrated.

References

[1] Bai, Z.D. Miao, B.Q. and Pan, G.M. On asymptotics of eigenvectors of large sample covariance matrix. Ann. Probab. To appear.

[2] Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann. Probab.* **32** 553-605.

[3] Silverstein, J. W. (1990). Weak convergence of random functions defined by the eigenvectors of sample convariance matrices. *Ann. Probab.* **18** 1174-1193.

Session 15

Time Series

Principal Component Analysis of Dependent Data

Paulo Canas Rodrigues, paulocanas@fct.unl.pt
Department of Mathematics and CMA. Faculty of Science and Technology, Nova University of Lisbon, Monte de Caparica, 2829-516 Caparica, Portugal.

Keywords: Principal Component Analysis. Time Series. SSA. MSSA.
AMS: 62H25, 62M15

Abstract

In many of the statistical methods, and in particular in Principal Component Analysis (PCA), it is assumed that all n observations $x_1, x_2, \ldots, x_n$ are independent. The aim of this work is to analyze the consequences of the application of PCA to observations that do not verify this hypothesis.

Skinner [5] and Konishi [2], among others, state that the type of dependence in the observations is varied. Being time series the most frequent situation in practice, we present the Singular Spectrum Analysis (SSA), Golyandina [1], and the Multichannel Singular Spectrum Analysis (MSSA), Oliveira [4] or Kim [3] as methods based in PCA.

This work is divided in two parts. In the first we present a description of SSA, which uses the auto-covariance matrix of a time series with only one variable and one application. In the second part we present the multivariate case, the MSSA, together with another application.

References

[1] Golyandina, N., V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Séries Structure: SSA and Related Thechniques*, Chapman & hall, 2001.

[2] Konishi, S., AND C. R. Rau, “Principal component analysis for multivariate familial data,” *79. Biometrika*, pp. 631–641, 1992.

[3] Kim, K. Y., AND Q. Wu, “A Comparison Study of EOF Techniques: Analysis of Nonstationary Data with Periodic Statistics,” *12. Journal of Climate*, pp. 185–199, 1998.

[4] Oliveira, I., *Correlated Data in Multivariate Analysis*, University of Aberdeen, 2003.

[5] Skinner, C. J., D. J. Holmes, AND T. M. F. Smith, “The effect of sample design on principal component analysis,” *81. J. Amer. Statist. Assoc.*, pp. 789–798, 1986.

Comparison of prediction quality of the best linear unbiased predictors in time series linear regression models.

Martina Hančová, martina.hancova@upjs.sk
Institute of Mathematics, P. J. Šafárik University in Košice
Jesenná 5, 040 01 Košice, Slovakia

Keywords: Time series. Finite discrete spectrum linear regression models. Best linear unbiased prediction. Mean squared error of a predictor. Löwner partial ordering.
AMS: 62M10, 62M20

Abstract

Using a matrix-block version of the mean squared error of the best linear unbiased predictor (BLUP) and the Löwner partial order relation for symmetric matrices, we derive and present our results dealing with a comparison of prediction quality of BLUP in different types of time series linear regression models which belong to a general class of the so-called finite discrete spectrum linear regression models (FDSLRM). Some of the results are completely new and some of them are generalization of recent results achieved in special orthogonal cases of FDLSRM.
Acknowledgements: This work was supported by the grant 1/3001/06 of the Slovak Scientific Grant Agency (VEGA) and the grant 25/07-08 of VVGS at P. J. Šafárik University in Košice.

Session 16

Statistical Applications II

Recent Advances in Palaeoclimate Reconstruction

Andrew Parnell, andrew.parnell@tcd.ie
Department of Statistics, Trinity College, Dublin, Ireland.

Keywords: Bayesian statistics; Spatial Models; Paleostatistics
AMS:

Abstract

This talk will concentrate on some of the developments since Haslett et al. [1]. The premise is that climate change affects pollen distribution; fossil pollen from lake sediment thus gives information on historical climate. By building a training data set which links modern pollen and modern climate, we can calibrate fossil pollen into historical climate. The Bayesian formulation of this problem contains many interesting facets, including spatial modelling in high dimensions, radiocarbon dating of depth chronologies, and some aspects of zero-inflation. In particular, the talk will discuss how smoothness in climate change is modelled, and extensions which aim to reconstruct the entire climate of Western Europe.

References

[1] J. Haslett, J., Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S.P., Alllen, J.R.M., Huntley, B., Mitchell, F. (2006). *Bayesian Palaeoclimate Reconstruction*. Journal of the Royal Statistical Society, Series A, 169, 3, 395-438

On spatial extremes: with application to a rainfall problem

Chen Zhou, zhou@few.eur.nl
Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000DR Rotterdam, The Netherlands

Keywords: spatial extremes, max-stable process, areal reduction factor.
AMS: 60G70; 62G32; 62M30;

Abstract

We consider daily rainfall observations at 32 stations in the province of North Holland (The Netherlands) during 30 years. Let Q be the *total* rainfall in this area on one day. An important question is: what is the amount of rainfall Q that is exceeded once in 100 years? This is clearly a problem belonging to extreme value theory. Also it is a genuinely spatial problem.

Recently, a theory of extremes of continuous stochastic processes has been developed. Using the ideas of that theory and much computer power (simulations) we have been able to come up with a reasonable answer to the question above.

Session 17

Theory of Codes

Probabilistic models in quasigroup error-detecting codes

Nataša Ilievska, natasha@ii.edu.mk
Institute of Informatics, Faculty of Natural Sciences and Mathematics, P.O.Box 162, Skopje, Republic of Macedonia

Keywords: probabilistic models, error-detecting codes, quasigroup.
AMS: 94B70

Abstract

In my future work I plain to investigate error-detecting codes and probabilistic models in them. Error-detecting codes are used to detect errors when messages are transmitted through a noisy communication channel. We propose error-detecting codes using quasigroup operation $*$ on set $A = \{0, 1, 2, 3\}$. In order to detect errors, we extend an input message $a_1a_2 \dots a_n$ to block $a_1a_2 \dots a_nd_1d_2 \dots d_n$, where $d_i = a_i * a_{(i+1) \ mod \ n}$, $i = 1, 2, \dots, n$. We calculate an approximate formula which gives the probability that there will be errors which will not be detected.
Acknowledgements: These problem was formulated by professor Verica Bakeva, Institute of Informatics, Faculty of Natural Sciences and Mathematics, Skopje.

References

[1] Markovski, S., Bakeva, V. (2005) *Error-detecting codes with cyclically defined redundancy*, Procedings of 3-th Congress of Mathematicians of Macedonia, (in print)

[2] Markovski,S., Bakeva,V. (2001) *Quasigroup Based Stream Error Correcting Codes*, Proceedings of the 2nd CIIT, Molika, R.Macedonia, p.14-19

[3] Vanstone, S.A., van Ooschot, P.S. (1989) *An Introduction to Error Correcting Codes with Applications*, Kluwer Academic Publishers, Boston.

[4] Hill, R. (1986) *A First Course of Coding Theory*, Clarendon Press, Oxford.

Random codes

Aleksandra Popovska, aleksandrap@ii.edu.mk
Institute of Informatics, Faculty of Natural Sciences and Mathematics, P.O.Box 162, Skopje, Republic of Macedonia

Keywords: random codes, probabilistic models.
AMS: 94B60

Abstract

The non-constructive proof of the noisy-channel coding theorem shows that good block codes exist for any noisy channel, and moreover that nearly all block codes are good. However, writing down an explicit and practical encoder and decoder that are as good as proved by Shannon in his seminal work A Mathematical Theory of Communication is still an unsolved problem. Recently, it has been recognized that two classes of codes, namely turbo codes and low-density parity-check (LDPC) codes, perform at rates extremely close to the Shannon limit. Turbo and LDPC codes are based on a similar philosophy: constrained random code, described by some fixed parameters plus randomness, decoded using iterative algorithms or message passing decoders. My plains for future work are investigations to probabilistic models in coding theory, especially in random codes.

References

[1] Gligoroski, D., Markovski, S., Kocarev, Lj. (2005) *Efficiently Decodable Almost Random Codes*, Submitted to IEEE Information Theory

[2] Barg,A., Forney,D. (2002) *Random Codes: Minimum Distances and Error Exponents*, IEEE Transactions on Information Theory, VOL. 48, NO 9, p.2568-2573

[3] MacKay,D.,(2003) *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press 2003,

[4] Vanstone, S.A., van Ooschot, P.S. (1989) *An Introduction to Error Correcting Codes with Applications*, Kluwer Academic Publishers, Boston.

Index

CiεM

UC
UNIVERSIDAD DE CANTABRIA

UEX
UNIVERSIDAD DE EXTREMADURA

RESURGO
EADEM
MUTATA

SEIO.ES

MINISTERIO
DE EDUCACIÓN
Y CIENCIA

i-math ingenio mathematica
consolider ingenio2010