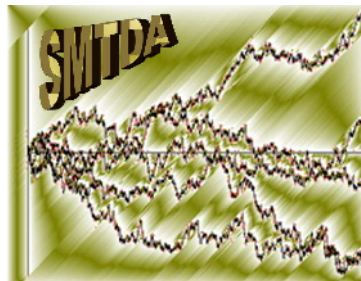# SMTDA 2010

Book of Abstracts

Stochastic Modeling Techniques and Data Analysis

International Conference

Editor

Christos H. Skiadas

June 8 - 11, 2010

Chania Crete Greece

MAICh

Prefecture of Chania

Swets

Introduction

Stochastic Modeling Techniques and Data Analysis
International Conference (SMTDA 2010)

Chania, Crete Greece, June 8 - 11, 2010

It is our pleasure to welcome the guests, participants and contributors to the International Conference (SMTDA 2010) on Stochastic Modeling Techniques and Data Analysis.

The main goal of SMTDA 2010 is to promote new methods and techniques for analyzing data, in fields like stochastic modeling, optimization techniques, statistical methods and inference, data mining and knowledge systems, computing-aided decision supports, neural networks and chaotic data analysis.

SMTDA aims at bringing together people from both stochastic and data analysis areas. Special attention is given to applications or to new theoretical results having potential of solving real life problems.

SMTDA2010 International Conference focus in expanding the development of the theories, the methods and the empirical data and computer techniques, and the best theoretical achievements of the Stochastic Modeling Techniques and Data Analysis field, bringing together various working groups for exchanging views and reporting research findings.
We thank all the contributors to the success of this conference and especially the authors of this *Book of Abstracts* of SMTDA 2010.

Chania, May 2010

Christos H. Skiadas
Conference Chair

Scientific Committee

C. H. Skiadas, Technical University of Crete, Chania Crete Greece (Chair)
A. Andronov, Transport and Telecommunications Institute, Riga, Latvia
Y. Bakouros, Industrial Management Lab., Univ. of Western Macedonia, Greece
N. Balakrishnan, McMaster University, Canada
M. H. Bacelar Nicolau, University of Lisbon, Portugal
H. W. Block, Department of Statistics, University of Pittsburgh, USA
M. Brown, The City College, CUNY, USA
E. Bulinskaya, Moscow State University, Russia
J. Caiado, Centre for Applied Mathematics and Economics, Technical University of Lisbon, Portugal
E. Canuto, Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy
T. Chatzipantelis, Aristotle University of Thessaloniki, Greece
M. Christodoulou, Technical University of Crete, Chania, Greece
I. Dimotikalis, Technological Educational Institute of Crete, Greece
D. Donchev, St. Kliment Ohridsky University, Bulgaria
A. Giovanis, Technological Educational Institute of Athens, Greece
J. Glaz, University of Connecticut, Stors, USA
G. Govaert, Universite de Technologie de Compiegne, France
A. Guenoche, University of Marseille, France
Y. Guermeur, LORIA-CNRS, France
S. Haberman, Cass Business School, City University, London, UK
D. Ho, IBM, Paris, France
P. Imkeller, Humboldt-Universitaet zu Berlin, Germany
J. Janssen, EURIA et ENST-Bretagne, Brest, France and JACAN, Tubize, Belgium
U. Kamps, Institut fur Statistik und Wirtschaftsmathematik, RWTH Aachen, Germany
A. Karagrigoriou, University of Cyprus, Cyprus
A. Katsirikou, University of Piraeus, Greece (Secretary)
W. Klonowski, Lab. Biosignal Analysis Fundamentals, Polish Academy of Sciences, Poland
A. Kohatsu-Higa, Osaka University, Osaka, Japan
G. E. Kokolakis, National Technical University of Athens, Greece
T. Kollo, Institute of Mathematical Statistics, Tartu, Estonia
N. Komninos, URENIO, Aristotle University of Thessaloniki, Greece
D. G. Konstantinides, Dept. of Statistics and Actuarial Science. University of the Aegean, Greece
E. J. Kontoghiorghes, Univ. of Cyprus and Queen Mary, Univ. of London, UK
S. Lallich, University Lyon 2, France
L. Lebart, CNRS, ENST, Paris, France
M.-L. Ting Lee, University of Maryland, USA
C. Lefevre, ULB, Bruxelles, Belgium
P. Lenca, Telecom Bretagne, France
N. Limnios, University de Technologie de Compiene, France
Bo H. Lindquvist, Norvegian Institute of Technology, Norway
G. MacKenzie, Univerity of Limerick, Ireland
R. Manca, University La Sapiensa, Roma, Italy
L. Markus, Eotvos Lorand University, Hungary
S. McClean, University of Ulster at Coleraine, Northern Ireland
L. Minkova, Department of Probability, Operations Research and Statistics, Univ. of Sofia, Bulgaria
K. Mosler, University of Koeln, Germany
T. Papaioannou, Universities of Pireaus and Ioannina, Greece
L. Sakalauskas, Institute of Mathematics and Informatics, Vilnius, Lithuania
G. Saporta, CNAM, Paris, France
W. Sandmann, Department of Mathematics, Clausthal University of Technology, Germany
J. M. Sarabia, Department of Economics, University of Cantabria, Spain
P. Sirirangsi, Chulalongkorn University, Thailand
C. Skiadas, Hanover College, Indiana, USA
R. Szekli, University of Wroclaw, Poland
T. Takine, Osaka University, Japan
P. Taylor, University of Melbourne, Australia
M. J. Valderrama, Universidad de Granada, Spain
P.-C.G. Vassiliou, Aristotle University of Thessaloniki, Greece
W. Wefelmeyer, Mathematical Institute, University of Cologne, Germany
V. Zaiats, Escola Politecnica Superior, Universitat de Vic, Barcelona, Spain
K. Zografos, Department of Mathematics, University of Ioannina, Greece

Keynote Talks

Narayanaswamy Balakrishnan
Department of Mathematics and Statistics
McMaster University, Hamilton, Ontario
Canada
Some Cure Rate Models and Associated Inference and Application to Cutaneous Melanoma Data

Erricos John Kontoghiorghes
University of Cyprus and Queen Mary, University of London, UK
Efficient algorithms for computing the best subset regression model

Rémi Léandre
Institut de Mathématiques de Bourgogne
Université de Bourgogne , Dijon
France
Wentzel-Freidlin estimates in semi-group theory

Nozer D. Singpurwalla
Institute for Reliability and Risk Analysis
Department of Statistics, The George Washington University
Washington, D.C., USA
Network Routing in a Dynamic Environment
Related Book: *Reliability and Risk: A Bayesian Perspective, Wiley, 2006*

Contents

Long memory based approximation of filtering in non linear switching systems

Noufel Abbassi and Wojciech Pieczynski

CITI Department, Institut Telecom, Telecom Sudparis, Evry, France
wojciech.pieczynski@it-sudparis.eu

Let $X_1^N = (X_1,...,X_N)$, $R_1^N = (R_1,...,R_N)$ and $Y_1^N = (Y_1,...,Y_N)$ be three random processes. Each X_n and Y_n take their values from $\mathbb{R}$, while each R_n takes its values from a finite set $\Omega = \{\omega_1,...,\omega_K\}$. The sequences X_1^N and R_1^N are hidden and the sequence Y_1^N is observed. We deal the classical problem of filtering, which consists of computation of $p(r_{n+1}|y_1^{n+1})$ and $E[X_{n+1}|r_{n+1}, y_1^{n+1})$ with a reasonable complexity. We consider the following partly non linear model:

R is a Markov chain;

$X_{n+1} = F_n(R_{n+1})X_n + H_n(R_{n+1})W_{n+1}$; and

$Y_n = G_n(R_n, X_n) + K(R_n)Z_n$.

The aim of the paper is to propose a partially unsupervised filtering method based on the recent model proposed in [3], and to compare its efficiency to the efficiency of classical models based, partly [2, 4, 5] or entirely [1], on the particle filtering. We present a rich study showing the interest of our model and related filtering method, which thus appear as a viable alternative to the classical methods.

References
1. Andrieu, C., Davy, M. and Doucet, A. Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions, *IEEE Trans. on Signal Processing*, **51**, 7, 1762-1770 (2003).
2. Kaleh, G. K. and Vallet, R. Joint parameter estimation and symbol detection for linear or nonlinear unknown channels. IEEE Trans. on Communications., 42, 7, 2406-13 (1994).
3. Pieczynski, W., Abbassi, N. and and Ben Mabrouk, M., Exact Filtering and Smoothing in Markov Switching Systems Hidden with Gaussian Long Memory Noise, XIII International Conference Applied Stochastic Models and Data Analysis, (ASMDA 2009), June 30- July 3, Vilnius, Lithuania (2009).
4. Tanizaki, H. *Nonlinear Filters, Estimation and Applications*. Springer, Berlin, 2nd edition (1996).
5. Wan, E. A. and Van der Merwe, R., 'The unscented Kalman filter for nonlinear estimation,' in Proc. IEEE Symp. Adaptative Systems for Signal Proc., Comm. And Control (AS-SPCC), (Lake Louise Alberta, Canada), 153-158 (2000).

Approximate Algorithm for Unconstrained (Un) Weighted Two-dimensional Guillotine Cutting Problems

Ait Abdesselam Maya and Ouafi Rachid

Laboratoire LAID3, Departement of Operations Research, Faculty of Mathemathics, USTHB, Algiers, Algeria, mayaaitabdesselam@yahoo.fr

The unconstrained two-dimensional guillotine cutting (U_TDC) problem consists of determining a cutting pattern of a set of n small rectangular pieces on an initial rectangular support of length L and width W, as to maximize the sum of the profits of the pieces to be cut. Each piece i, is characterized by a length li , a width wi and a profit vi(or weight) . In this paper we propose an approximate algorithm for solving both unweighted and weighted unconstrained two-dimensional guillotine cutting problems (UU_TDC and UW_TDC). The original problem is reduced to a series of single bounded knapsack problems and solved by applying a dynamic programming procedure. We evaluate the performance of the proposed method on instances taken in the literature.
Key Words: guillotine cutting, dynamic programming, knapsack problem

Regression Spline Modelling of the Effects of Past Values of Time-Varying Prognostic Factors on Survival

Michal Abrahamowicz and Marie-Pierre Sylvestre

Department of Epidemiology & Biostatistics, McGill University, Montreal, Quebec, Canada
michal.abrahamowicz@mcgill.ca
Lunefeld Research Institute, Toronto, Ontario, Canada, sylvestre@lunenfeld.ca

Time-dependent covariates (TDC) are increasingly used in survival analysis to model the effects of prognostic and/or risk factors, as well as treatments, whose values change during the follow-up. Modelling of TDC requires specifying the relationship between the hazard and the entire vector of its past values, rather than a scalar, and accounting for likely cumulative effects of the past values. This requires assigning differential weights to TDC values observed at different times in the past, but such weights are generally unknown and have to be estimated from the data at hand. Furthermore, the form of the dose-response relationship between the TDC value and the hazard is also typically unknown and has to be estimated. We extend our recent, simpler method [1] to model the cumulative effect of TDC, with simultaneous estimation of (i) weight function, as in [1], and (ii) possibly non-linear dose response curve. The weighed cumulative exposure (WCE) effect at time τ, is a function of the time-dependent vector of past TDC x(t), at $0 < t < \tau$:

$$WCE(\tau|x(t), t < \tau) = \sum w(\tau - t) \cdot s[x(t)] \quad (1)$$

where $w(\tau - t)$ is a weight function and $s[x(t)]$ a dose-response function. Both $w(\tau - t)$ and $s[x(t)]$ are modeled using low-dimension cubic regression splines. The estimated WCE is then included as a TDC in the Cox's PH model. The model (1) is estimated using iterative 2-step alternating conditional estimation algorithm, which alternates between estimation of $w(\tau - t)$ and $s[x(t)]$. The estimates of the two functions are evaluated in simulations. The model is then applied to re-assess the cumulative effects of blood pressure, measured at different times in the past, on cardiovascular risks.

Keywords: Survival, Cumulative effects, Splines

References

1. Sylvestre MP, Abrahamowicz M. Flexible modelling of the cumulative effects of time-dependent exposures on the hazard. Stat Med 2009;28:3437-53.

Telecenters through Wireless Technology near a Base camp of Mt. Everest of Nepal

Bimal Acharya

PhD Candidate Tribhuvan University), Fellow IETE, Master of Engineering (Telecom, AIT-Bangkok), B.Sc. Engineering (EEE, BUET, Dhaka), bimal.acharya@ntc.net.np

Wireless communications in Nepal has expanded greatly in last decade and the number of the wireless users has grown to more than 1500 thousand. However, the distributions are more concentrated in major cities. Thus there are double challenges we are facing namely how to take telecom revolution to the rural areas especially to the mountainous villages and tourism area. This article focuses how to provide voice, email and internet services to the mountainous tourism villages: near a Base Camp of Mount Everest, the summit of world. Satellite link is presented as only one of the quick and cost effective option for providing backhaul link to this proposed project. Wireless LAN (Wi-Fi) is chosen as technology for connecting different villages or business houses or trekking camps of that area. These Telecenters is expected not only to revolutionize the system of social participation of local people by internet and communications but also attracts the foreigners for trekking. Social awareness programs including education of rural mass can be achieved in most cost effective way in the shortest possible time to an extremely larger group of rural people. Installation of Telecenters certainly improves the life styles of local people as well as facilitates the tourists.

Key words: Telecenters, Backhaul, Information Communication Technology (ICT), WiFi, Satellite Communications, LAN, Cyber Cafe, Broadband.

Short Resume:

Er. Bimal Acharya is a life member of NEA. He is also life fellow/member of The Institution of Electronics and Telecommunications Engineers (IETE) India and is acting as Honorary Secretary of IETE Nepal since 2006 as well as other premier professional organizations. He is a PhD candidate of Tribhuvan University, Nepal. He holds:

- Master of Engineering in Telecommunications from Asian Institute of Technology, Thailand in 1999
- Bachelor of Science and Engineering (Electrical and Electronics), Bangladesh University of Engineering and Technology, Dhaka, Bangladesh 1994.
- Er. Acharya had served the Nepal for more than 15 years as an engineer and senior engineer, in various capacities including, In-charge of VSAT unit of Rural Services Directorate, Nepal, Member of different technical specifications preparation and evaluation committees. He had already visited more than 71 districts of Nepal during the survey, planning, installation and Operation and Maintenance of rural telecommunications of Nepal. Academically, he is a popular part time faculty member of Kathmandu University, Institute of Engineering, Pokhara University, Purbanchal University. He had taught Wireless communications, Satellite communications, Telecommunications, Digital Communications, Communications engineering for Master level and different communications courses for Undergraduate levels of different colleges.

Performance evaluation of IP networks with differentiated services

S. Adjabi, K. Adel-Aissanou, N. Saidani and A. Laouar

Laboratory LAMOS, University of Bejaia, Bejaia, Algeria, adjabi@hotmail.com

The aim of this study is the mathematical modelling of the architecture with differentiation of services (DiffServ) as it is defined by the IETF workgroup with an M/G/1/N queue with multiple vacations and exhaustive service. We are interested by the evaluation of the performances of the EF (Expedited Forwarding) class of the core routers which contains the most critical packages (voice, video). For the validation of the analytical chosen model, a simulation of a simple network under DiffServ was realized with NS2. This modelling (with M/G/1/N queues) serves for the sizing of the network parameters: debit and size of buffer according to the load of the network.

Key Words: IP networks, DiffServ (differentiated services), M/G/1/N queue with multiple vacation

Queueing models with periodic input processes.

Larisa G. Afanasieva and Elena E. Bashtova

Department of Theory of Probabilities, Moscow State University, Russia,
afanas@mech.math.msu.su
Department of Theory of Probabilities, Moscow State University, Russia,
bashtovaelena@rambler.ru

The motivation for the investigation of queueing systems with periodic flows comes from numerous applications. Many applied problems from such areas as reliability, inventory theory, computer science, transport networks and so on, can be embedded into waiting problems for suitably selected input flow with time-dependent periodic intensity and model construction. These problems are also of theoretical interest. They are often associated with the analysis of differential and integro-differential equations with periodic coefficients.

In this talk we confine ourselves to two problems: the determination of conditions of the existence of a limit periodic distribution and the asymptotical analysis of the queue under various assumptions.

Queues with non-homogeneous Poisson arrivals were studied by many authors (see for ex. Takach L., Reich E., Hasofer A.M., Harrison J.M. and Lemoin B.J., Rolsky T.). One of the main new elements in the present study is that we consider systems with a general periodic input X(t). We recall that a stochastic flow X(t) is periodic with period T if the finite-dimensional distributions of its increments are T-time-shift invariant.

Stochastic boundedness conditions for systems with such input flows are given. Under some additional assumptions the existence of the limit periodic distribution is determined. Then we focus our attention on the asymptotic properties of this distribution. Especially the heavy and light situation are considered. Diffusion approximation type theorems for trajectories of the processes that describe the system under consideration (waiting time, number of customers) are proved. Our investigations are based on the construction of majorizing systems, the use of results from renewal theory and the application of asymptotical analysis of queueing models with stationary control sequence. To explain the last approach, denote by $\{a_k\}_{k=1}^{\infty}$ a sequence of interarrival times for a periodic flow. Under some natural assumptions a sequence $\{a_{n+k}\}_{k=1}^{\infty}$ weakly converges as $n \to \infty$ to a stationary one. This convergence allows us to apply traditional asymptotical methods of queueing theory to the analysis of periodic models.

Queueing Systems in Regenerative Random Environment

Afanasyeva L.G., Belorusov T.N.

Mechanics and Mathematics Department, Moscow State University, Moscow, Russia
afanas@mech.math.msu.su, timophey.belorusov@gmail.com

For queueing systems in a random environment conditions of limit regimes existence are found. A stochastic process $\{X(t), t \geq 0\}$ with nondecreasing left-continuous trajectories is named a stochastic flow. Assume $\{U(t), t \geq 0\}$ is a regenerative process with consequent moments of regeneration $\{\theta_i\}$, $i \geq 1$ and periods of regeneration $\tau_i = \theta_i - \theta_{i-1}$, where $i \geq 1$ and $\theta_0 = 0$.

Definition 1. A stochastic flow X(t) is said to be functioning in the random environment U(t) if the sequence $\{\chi_i\} = \{X(\theta_{i-1}+t) - X(\theta_{i-1}), 0 \leq t < \tau_i, \tau_i\}$, $i \geq 1$ consists of independent random elements, and $\{\chi_i\}$, $i \geq 2$ are identically distributed.

Denote $\xi_i = E[X(\theta_i) - X(\theta_{i-1})]$ and $\mu = E\tau_2$, $a = E\xi_2$. If $\mu < \infty$, $a < \infty$ and $P(\theta_1 < \infty) = 1$, then the stochastic flow has an intensity $\lim_{t \to \infty} X(t)/t = a/\mu$, a.s.

The majority of flows considered in queueing theory falls within this class. They also find useful applications in inventory theory, reliability theory, transportation nets, etc. Keeping in mind applications it is natural to consider service process that depends on the random environment as well. So let Y(t) be the amount of work the server can perform during the time interval [0, t]. We say that a queueing system works in a random environment U(t) if at least one of the processes X(t) or Y(t) is functioning in this environment in the sense of Definition 1.

To illustrate the concept we introduce several models. Here we give two of them. First, consider a single-server system with input X(t) and service described by the process Y(t) in the random environment U(t). To avoid formal difficulties we suppose that the distribution of τ_2 has an absolutely continuous component. We define the traffic coefficient $\rho = \beta a^{-1}$ with $\beta = E[Y(\theta_2) - Y(\theta_1)]$. Let W(t) be a workload process. Under these conditions we prove that the process W(t) has a proper limit distribution iff $\rho < 1$.

Second, suppose there is a sequence $\{f_j, j=0,1,\ldots\}$, where f_j is in (0,1]. A new customer encountering j other customers in the system stays for service with probability f_j and gets rejection with probability $1 - f_j$. Service times are represented by independent and identically distributed random variables with an arbitrary distribution function and finite mean. Input X(t) is functioning in the random environment U(t). We prove that the process W(t) has a proper limit distribution if $\limsup_{t \to \infty} f_j < (\lambda b)^{-1}$, and W(t) is stochastically unbounded if $\liminf_{t \to \infty} f_j > (\lambda b)^{-1}$. Some examples that could be interesting in applications (systems with Doubly Stochastic Poisson Flow, Semi-Markov Flow, transportation nets, etc.) are also given.

Customer Centric Approach Using Data Mining Techniques

Vasilis Aggelis and Georgia Peleki

Business Analysis and Customer Utilization Team, Piraeus Bank SA, Athens, Greece
aggelisv@piraeusbank.gr, pelekig@piraeusbank.gr

Nowadays the majority of markets experience the global financial crisis. Is such special situation financial institutions have to adopt and deploy smart models. It is common sense that banks can no longer stay competitive just by squeezing more efficiencies and productivity gains from operational applications. Banks have to spend time in order to know their customers and their needs. Taking advantage all available, internal and external, information, banks should target to increase market share, effectiveness of marketing campaigns, enter new markets, and ultimately deliver more products and services that would be suitable for customers needs. K.R.I.O.S. model and data mining techniques are the keys to providing the right information for implementing the aforementioned business actions. In this paper we present some real case studies, concerning customer segmentation, customer development plan and product development. Those case studies are customer centric and show that if we have a steady and business well-founded target then data could lead us in realistic solutions.
Keywords: Data Mining, Decision support, Marketing campaigns, Customer Segmentation

ABOUT STOCHASTIC OPTIMAL CONTROL PROBLEM OF SWITCHING SYSTEM WITH DELAY

Aghayeva Ch. and Morali N.

Institute of Cybernetics, Baku, Azerbaijan
Email: agayeva.cherkez@yasar.edu.tr
Department of Statistics, Yasar University
Izmir, Turkey
Email: nilgun.morali@yasar.edu.tr

Let $(\Omega, F^l, P), l = \overline{1,r}$ be a probability spaces with filtration $\left\{ F_t^l, t \in [t_{l-1}, t_l], l = \overline{1,r} \right\}$, $0 = t_0 < t_1 < ... < t_r \leq T$. Let $w_t^1, w_t^2, ..., w_t^r$ are independent Wiener processes, $F_t^l = \overline{\sigma}(w_q^l, t_{l-1} \leq q \leq t \leq t_l), l = \overline{1,r}$. $L_{F^l}^2(a, b; R^n)$ is the space of all predictable processes such as $E\int_a^b |x_t(\omega)|^2 dt < +\infty$. $R^{m \times n}$ is the space of linear transformations from R^m to R^n.

Let $O_l \subset R^{n_l}, Q_l \subset R^{m_l}, l = \overline{1,r}$, be open sets, R^n be an n- dimensional Euclidean space, $\Lambda_l, l = \overline{1,r}$ be the set of continuous functions, $T = [0, T]$ be a finite interval.
Consider the following stochastic control system:

$$dx_t^l = g^l\left(x_t^l, x_{t-h_l}^l, u_t^l, t\right)dt + f^l(t)dw_t^l \quad t \in (t_{l-1}, t_l], l = \overline{1,r} \text{ ,(1)}$$

$$x_t^{l+1} = K^l(t) + \Phi^l\left(x_t^l, t\right), t \in [t_l - h_l, t_l], l = \overline{1, r-1};$$
$$x_{t_0}^1 = K^0(t), t \in [t_0 - h_0, t_0] \tag{2}$$

$$u_t^l \in U_{\partial}^l \equiv \left\{ u^l(\cdot, \cdot) \in L_F^2\left(t_{l-1}, t_l; R^m\right) | u^l(t, \cdot) \in U^l \subset R^m, l = \overline{1,r} \quad a.c. \right\} \tag{3}$$

where $U^l, l = \overline{1,r}$ are non-empty bounded sets, $K^l(\cdot), l = \overline{1,r}$ are non-random functions and $h_l \geq 0, l = \overline{1,r}$.

The problem is concluded to find the controls $u^1, u^2, ..., u^r$, trajectories $x^1, x^2, ..., x^r$ and the switching law $t_1, t_2, ..., t_r$ which are minimize the cost functional :

$$J(u) = \sum_{l=1}^{r} J^l\left(u^l\right) = \sum_{l=1}^{r} E\left[\varphi^l\left(x_{t_l}^l\right) + \int_{t_{l-1}}^{t_l} p^l\left(x_t^l, u_t^l, t\right)dt\right] \qquad (4)$$

which determined on the decisions of the system (1)-(3), generated

by all admissible controls $U = U^1 \times U^2 \times ... \times U^r$ at conditions:

$q^l\left(x_{t_r}^r, t_r\right) = 0, \ l = \overline{1, r}$; (5).

The following result that is a necessary condition of optimality for problem (1)-(5) has been obtained.

Theorem. Suppose that, $\pi^r = (t_0, t_1, t_r, x_t^1, x_t^2, ..., x_t^r, u^1, u^2, ..., u^r)$ is a optimal solution of

problem (1)-(5) and random processes $(\psi_t^l, \beta_t^l) \in L_{F^l}^2(t_{l-1}, t_l; R^{n_l}) \times L_{F^l}^2(t_{l-1}, t_l; R^{n_l x n_l})$ which are

the solutions of the following adjoint equations:

$$\begin{cases} d\psi_t^l = -\left[H_x^l(\psi_t^l, x_t^l, u_t^l, t) + H_y^l(\psi_{t+h_l}^l, x_{t+h_l}^l x_t^l, u_t^l, t)\right]dt + \beta_t^l dw_t^l, \ t_{l-1} \le t < t_l - h_l, \\ d\psi_t^l = -H_x^l(\psi_t^l, x_t^l, u_t^l, t)dt + \beta_t^l dw_t^l, \ t_{l-1} - h_l \le t < t_l, \ l = \overline{1, r}, \\ \psi_{t_l}^l = -\varphi_x^l(x_{t_l}^l) + \psi_{t_{l+1}}^l \Phi_x^l(x_{t_l}^l, t_l), l = 1, r-1, \\ \psi_{t_r}^r = -\lambda_0 \varphi_x^r(x_{t_r}^r) - \sum_{l=1}^{r} \lambda_l q_x^l(x_{t_r}^r, t_r), \end{cases}$$

where $\left(\lambda_0, \lambda_1, ..., \lambda_r\right)$ is a non-zero vector .

Then following conditions are fulfilled :

a) $\forall \tilde{u}^l \in U^l, l = 1, r,$ the following maximum principle:

$H^l(\psi_\theta^l, x_\theta^l, \tilde{u}^l, \theta) - H^l(\psi_\theta^l, x_\theta^l, u_\theta^l, \theta) \le 0,$ a.e. $\theta \in [t_{l-1}, t_l]$,

b) transversality conditions:

$$\sum_{l=1}^{r} \lambda_l q_t^l(x_{t_r}^r, t_r) - \sum_{l=1}^{r-1} \psi_{t_l}^{l+1} \Phi_t\left(x_{t_l}^l, t_l\right) = 0 \ .$$

References
1.Gikhman I.I.,.Skorokhod A.V. (1968), Stochastic Differential Equations , Kiyev: Naukova Dumka, 352 p. (Russian)Fick, A., *Poggendorff's Annalen*, **94**, 59-86 (1855).
2.Boukas El- Kebir(2006), Stochastic Switching Systems. Analysis and Design, Birkhauer, 405 p.
3.Bengea S.C., Raymond A.C.(2005), Optimal Control of Switching systems, // Automatica , 41, pp. 11-27.
4.Agayeva Ch., Abushov G. .(2009), Necessary condition of optimality for one nonlinear stochastic control system with variable structure, Abstracts of the International Conference "Stochastic Analysis and Random Dynamical systems",Ukrain pp.1-2.
5.Agayeva Ch., Abushov G. Necessary condition of optimality for stochastic control systems with variable structure, Proceedings of the 20-th International Conference "Continuous, Optimization and Knowledge – based technologies.", EurOPT' 2008, Lithuania, May 2008, pp. 77-81.

Identifying Patterns and Cybernetics Solutions to Global economic crises: A study of qualitative data analysis and modeling

Raman Kumar Agrawalla

Business Systems and Cybernetics Centre, Tata Consultancy Services, Hyderabad - 500 001, AP State, India, raman.a@tcs.com, raman_a@rediffmail.com

Crises, specifically economic crises, have been part and parcel of the global economy. The recent global economic recession, which is, of course, gradually subsiding, is a reality; set in since December 2007. There has been lot of discussions relating to it and regarding the causes and consequences of the global economic crisis. In economics academic fraternity,

recession is normally discussed in terms business cycles. Keynesian economics and Monetarism explains recessions as 'failure of some market to clear' and Real Business Cycle theories explain economic fluctuations as 'efficient response to exogenous changes in the real economic environment'. And of late, researchers in this area also base their discussions within the Dynamic Stochastic General Equilibrium (DSGE) framework. However, the present paper tries to discuss the topic in terms of concepts from Cybernetics, the science of systems control and regulation. Cybernetics is derived from the Greek kybernetes meaning 'steersman' or 'governor'. The term cybernetics was coined by Norbert Weiner, 1948 to refer to an area of study that covered 'the entire field of control and communication theory, whether in the machine or the animal'. After a decade, Pask, 1961 defined cybernetics as the study of 'how systems regulate themselves, reproduce themselves, evolve and learn'. The relevance and importance of cybernetics is supreme in this context if we see the honest confession by Ben Bernanke, the Chairman of the Federal Reserve, in September 2008, after the Lehman Brothers collapse, 'we have a total loss of control'. Through the present work, we will illustrate the methods to undertake qualitative data analysis and modeling in this subject of societal and business importance.

The present paper summarizes the history of economic theories in relation to the history of global recessions and attempts to review the stochastic modeling work in this area. On this back ground, we attempt in the paper to put forward the *patterns* that emerge in our qualitative analysis of global recessions and try to identify and present some *archetypes* of global recessions. Thus, the paper attempts to conceptualise a system to understand the global economic and financial crisis and draws upon cybernetic concepts as solutions towards better control and regulation of economic and financial systems. In particular, we try to invoke Ashby's theorem 'every good regulator of a system must be a model of that system' in our search for the cybernetic solutions. The point is how we can create a global economic system so that its *nervous system* withstands all short of perturbance and shocks. With this, the system can keep its ability to regulate and control intact; also will be able to assert itself to ward off any brewing financial crises, as and when needed. Also, how the industry and businesses can reorient themselves in such troubled times; the present paper also endeavours to throw some light to these aspects as well.

Stochastic Modeling of Web evolution

Efstathios Amarantidis[1], Ioannis Antoniou[2] and Michalis Vafopoulos[3]

[1]PhD candidate, Aristotle University of Thessaloniki, Mathematics Department, Graduate Program in Web,Veria, Greece, amartatis@yahoo.gr
[2]Professor, Aristotle University of Thessaloniki, Mathematics Department, Graduate Program in Web,Veria, Greece, iantonio@math.auth.gr
[3]Adjct Professor, Aristotle University of Thessaloniki, Mathematics Department, Graduate Program in Web,Veria, Greece, vafopoulos@webscience.gr

In the first twenty years of its existence, the Web has proven, to have had a fundamental and transformative impact on all facets of our society. While the Internet has been introduced 20 years earlier, the Web has been its "killer" application with more than 1.5 billion users worldwide accessing more than 1 trillion web pages (excluding those that cannot be indexed, the "Deep Web") (Googleblog 2008). Searching, social networking, video broadcasting, photo sharing, blogging and micro-blogging have become part of everyday life whilst the majority of software and business applications have migrated to the Web.

Today, the enormous impact, scale and dynamism of the Web in time and space demand more than our abilities to observe and measure its evolution process. Quantifying and understanding the Web lead to Web modeling, the backbone of Web Science research (Berners-Lee et al 2006). Web models should invest on Complexity (Antoniou et al 2000) beyond reductionism, linking structure to function and evolution (Prigogine 1999, Prigogine et al 2001, Meyers 2009). In this context, causality between events, temporal ordering of interactions and spatial distribution of Web components are becoming essential to addressing scientific questions at the Web techno-social system level.

The first steps towards understanding cyberspace involved measurements and statistics of the Internet traffic (Faloutsos et al 1999, Fabrikant et al 2002, Antoniou et al 2002 a, b, Antoniou et al2003). The self-similar feature of the Internet was also found in the Web through preferential attachment (Barabasi et al 1999 a, b, Albert et al 1999).

Network science being a useful mathematical framework to formulate the non-reducible interdependence of Complex Systems (Prigogine 1999, Prigogine et al 2001) recently led to significant results not only in Web graph statistics, but moreover in biology (Kitano 2002), economics (Easley et all 2010) and sociology (Liljeros et al 2001). These results initiated a new understanding of Complexity in nature (Newman et al 2006).

The statistical analysis of the Web graph led to four major findings (Bonato et al 2005): on-line property (the number of nodes and edges changes with time), power law degree distribution with an exponent bigger than 2, small world property (the diameter is much smaller than the order of the graph) and many dense bipartite subgraphs.

In the light of these findings Kouroupas, Koutsoupias, Papadimitriou and Sideri proposed an economic-inspired model of the Web (KKPS model thereafter) (Kouroupas et al 2005 a, b) which explains the scale-free behavior. Web evolution is modeled as the interaction of Documents, Users and Search Engines. The Users obtain satisfaction (Utility), when presented with some Documents by a Search Engine. The Users choose and endorse Documents with highest Utility and then the Search Engines improve their recommendations taking into account these endorsements, but not the dynamic interdependence of the Utility on the www state. Commenting on their results the authors have pointed out that (A) "more work is needed in order to define how the various parameters affect the exponent of the distribution" (of the in-degree of documents) and that (B) "increasing b (the number of endorsed documents) causes the efficiency of the algorithm to decrease. This is quite unexpected, since more user endorsements mean more complete information and more effective operation of the search engine. But the opposite happens: more endorsements per user seem to confuse the search engine."

The purpose of this paper is to address and clarify the issues (A), (B) arising within the KKPS modeling scheme (Kouroupas et al 2005 b), through analysis and simulations and to highlight future research developments in Web modeling.

Results and Discussion
1.   Concerning the dependence of the power-law exponent on the number α of recommended Documents by the Search Engine, the number k of topics and the number b of endorsed documents per User-Query, we found that the validity of the power law becomes less significant as b increases, both in the case α=b and in the case α≤b, confirming the results of Kouroupas et al [21]. Our simulations however, extended the investigation for different initial random distributions of the in-degree of Documents and for different values of α and b (Section 4).
2.   In the case α=b, Utility is useful only in terms of establishing compatibility between Utility Matrix and the Users-Queries and Documents bipartite graph, since all recommended Documents are endorsed according to the highest in-degree criterion.
3.   Concerning the origin of the power law distribution of the in-degree of Documents, two mechanisms are identified in the KKPS model:
•   Users-Queries endorse a small fraction of Documents presented (b).
•   Assuming a small fraction of poly-topic Documents, the algorithm creates a high number of endorsements for them.
The above mechanisms are not exhaustive for the real Web graph. Indexing algorithms, crawler's design, Documents structure and evolution should be examined as possible additional mechanisms for power law distribution.
4.   Concerning the dependence of the efficiency of the search algorithm (price of anarchy [21]) on the number α of recommended Documents by the Search Engine, the number k of topics and the number b of endorsed documents per User-Query we found that the efficiency of the algorithm increases, as the number α of recommended Documents by the Search Engine, the number k of topics and the number b of endorsed Documents per User-Query increase (Section 5). Our simulations confirmed the results of Kouroupas et al [21], except the dependence on the number b of endorsed documents per User-Query where they found that "increasing b causes the efficiency of the algorithm to decrease. This is quite unexpected, since more user endorsements mean more complete information and more effective operation of the search engine. But the opposite happens: more endorsements per

user seem to confuse the search engine." Therefore, in this case our result (Figure 7) confirmed their intuition but not their simulation.

5.   According to [21] "The endorsement mechanism does not need to be specified, as soon as it is observable by the Search Engine. For example, endorsing a Document may entail clicking it, or pointing a hyperlink to it." This hypothesis does not take into account the fundamental difference between clicking a link (browsing) and creating a hyperlink. Clicking a link during browsing is the "temporal" process called traffic of the Web sites [23]. Web traffic is observable by the website owner or administrator through the corresponding log file [24] and by third parties authorized (like search engine cookies which can trace clicking behavior [25] or malicious. On the contrary, creating a hyperlink results in a more "permanent" link between two Documents which is observable by all Users-Queries and Search Engines. Therefore, the KKPS algorithm actually examines the Web traffic and not the hyperlink structure of Documents which is the basis of the in-degree Search engine's algorithm.

6.   In this context, we remark that according to the published literature, Web traffic as well as Web content editing, are not taken into account in the algorithms of Search engines based on the in-degree (i.e. Pagerank [26]). These algorithms were built for Web 1.0 where Web content update and traffic monetization were not so significant. In the present Web 2.0 era with rapid change [27], the Web graph, content and traffic should be taken into account in efficient search algorithms. Therefore, birth-death processes for Documents and links and Web traffic should be introduced in Web models, combined with content update (Web 2.0) and semantic markup (Web 3.0 [28]) for Documents.

7.   The discrimination between Users and Queries could facilitate extensions of the KKPS model in order to incorporate teleportation (a direct visit to a Document which avoids Search Engines) to a Document, different types of Users and relevance feedback between Documents and Queries [29].

8.   In the KKPS model, Utility is defined to be time invariant linear function of R and D which by construction is not affecting the www state when $\alpha=b$. This is a first approximation which does not take into account the dynamic interdependence of the Utility on the www state. In reality, the evolution of the www state will change both R and D. A future extension of KKPS model should account for user behavior by incorporating Web browsing and editing preferences.

9.   Lastly, it would be very useful to offer deeper insight in the Web's business model by incorporating economic aspects in the KKPS model. This could be achieved by introducing valuation mechanisms for Web traffic and link structures and monetizing the search procedure (sponsored search [30]).

Keywords: Web science; Web modeling; scale-free networks; KKPS model; complex systems;

References

1. Albert R., Jeong H., Barabasi A.L., Diameter of the World-Wide Web, *Nature,* 401 (6749), 130 (1999).
2. Antoniou I., Reeve M., Stenning V., Information Society as a Complex System J., *Univ. Comp. Sci.* 6, 272-288 (2000).
3. Antoniou I., Ivanov Vi.V., Ivanov Valery V., Zrelov P.V., On the log-normal distribution of Network Traffic, *Physica D167*, 72-85 (2002).
4. Antoniou I., Ivanov V., Kalinovsky Yu., Kinetic model of Network traffic, *Physica A308*, 533-544 (2002).
5. Antoniou I., Ivanov Vi.V., Ivanov Valery V., Zrelov P. V., On a Statistical Model of Network Traffic, *Nucl. Instr. Meth. Phys. Res. 4502*, 768-771 (2003).
6. Barabasi A.-L., Albert R., Emergence of Scaling in Random Networks, in *Science 286* (5439), 509 (1999).
7. Barabasi A.-L., Albert R., Jeong H., Scale-free characteristics of random networks: the topology of the world-wide web, *Physica A 281*, 69-77 (2000).

Survival Analysis: an application to Italian data on public investments

Carlo Amati, Francisco Barbaro, Francesca Spagnolo, Fabio De Angelis, Maria Alessandra Guerrizio

Public Investments verification unit, Dept. for Development and Economic Cohesion, Italian Ministry of Economic Development, Rome, ITALY
mariaalessandra.guerrizio@tesoro.it

In Italy there has been a wide debate on the duration of implementation of public investment projects and more in general on the effectiveness of development policies: there is a common awareness that it often takes much too long to complete a planned infrastructure.

There are lots of monitoring systems that extensively collect data on the duration of the main stages of implementation of projects of public investments, together with information on structural features, such as location, sector, cost and implementing body which directly affect total length of projects.

Data refer to completed as well as to on-going projects.

In response to the need for knowledge and informed decisions on public investments, rather than common awareness and data repositories not thoroughly exploited, **survival models** have been applied to analyse these data.

Parametric models specific for each of the stages of implementation (design, procurement, works) have been defined and estimated, with two main aims:

- descriptive purposes:;
- predictive purposes.

Models' parameter estimates not only help identifying those factors that mostly affect the length of project's implementation, but also allow to forecast duration of on-going or future projects.

Parameter estimates are the engine of an interactive dashboard, VISTO, that given the project's characteristics, shows the predicted duration of a project's within a range of most probable values; thus, it can be used as a benchmarking tool for the ex-ante selection of projects during planning, as an early-warning system for on-going projects and for the ex-post assessment of the performance of completed projects.

Keywords: Survival Analysis, Parametric Models, Public Investments

Simple Non-Recurrent Flow and Its Applications in the Problems of Reliability, Storage and Queueing Theory

Alexander Andronov and Jelena Revzina

Transport and Telecommunication Institute, Lomonosova str. 1, LV-1019, Riga, Latvia
lora@mailbox.riga.lv, lena_revzina@tsi.lv

In this paper the non-recurrent flow of arrivals is considered in a case, where interarrival times $\{X_n\}$ correspond to the Markov chain with the continuous state space $R_+ = (0, \infty)$. The conditional probability density function of X_{n+1} given $\{X_n = z\}$ is determined by means of

$$q(x \mid z) = q(x \mid X_n = z) = \sum_{i=1}^{k} p_i(z) h_i(x), \quad z, x \in R_+,$$

where $\{p_1(z),..., p_k(z)\}$ is a probability distribution, $p_1(z) + ... + p_k(z) = 1$ for all $z \in R_+$; $\{h_1(x),..., h_k(x)\}$ is a family of probability density functions on R_+.

This flow is investigated for stationary case. One is considered as the Semi-Markov process $J(t)$ on the state set $\{1, …, k\}$. Main characteristics are considered: stationary distribution of J and interarrival times X, correlation and Kendall tau (τ) for adjacent intervals, and so on.

Further one is considered a Markovian system on which the described flow arrives. The system has a finite or countable set of (may be multidimensional) states N_+^r with $N_+ = \{0,\ 1,...,c\}$, $c \leq \infty$. Arrival moments of the flow transfer the system from a state $i \in N_+^r$ into other state $j \in N_+^r$ with probability

$$g_{i,j}^{(v)},\ g_{i,0}^{(v)} + g_{i,1}^{(v)} + ... + g_{i,c}^{(v)} = 1\,,$$

where v is the Semi-Markov component's J value immediately before the arrival.

Let 0 be the beginning of the new interval and w be the new value of J: $w = J\,(0+)$.

Between time moments of the new and the successive arrivals, a dynamic of the system is described by homogeneous Markov process $Y(t)$, $t > 0$, on the sate set N_+^r with transition probabilities

$$P_{i,j}^{(w)}(t) = P\big\{Y(t) = j \big| Y(0+) = i, L(0+) = w\big\},\ i, j \in N_+^r;\ w = 1,...,k;\ t \geq 0.$$

The stationary distribution and characteristics of the continuous-time process *(Y, J)* are considered.

Within the framework of the suggested model, various problems of reliability, storage and queues are considered.

Numerical results show that the dependence between interarrival times of the flow exercises greatly influences the efficiency characteristics of considered systems. Thus the study of stochastic models with dependencies is very important for applications.

Keywords: Semi-Markov process, Markov subordinator, stationary distribution

References
1. Kijima, M. *Markov Processes for Stochastic Modelling*: Chapman & Hall, London (1997).
2. Limnios, N. and Ouhbi, B. N*onparametric Estimation of Integral Functionals for Semi-Markov Processes with Application in Reliability in Statistical Models and Methods for Biomedical and Technical Systems*, (Vonta, F., Nikulin, M., Limnios, N., and Huber-Carol, C., eds.), Birkhauser, Boston-Basel-Berlin, 405 –418 (2008).
3. Pacheco, A., Tang, L.C., and Prabhu, U N. *Markov-Modulated process and Semiregenerative Phenomena:* World Scientific, New Jersey -London –Singapore (2009).

The Research of fractal characteristics electrocardiogram in a real time mode

Valery Antonov, Anatoly Kovalenko, Artem Zagaynov, Vu Van Quang

Saint-Petersburg State Technical University, Saint-Petersburg, Russia
antonovvi@mail.ru

The article presents the results of recent investigations into Holter monitoring of ECG, using non-linear analysis methods. It is shown that one of the most precise characteristics of the functional state of biological systems is the dynamical trend of correlation dimension and entropy.

On the basis of this it is suggested that a complex programming apparatus be created for calculating these characteristics on line. A similar programming product is being created now with the support of RFBR. The first results of the working programme, its adjustment, and further development, are also considered in the article.

Key Words: Holter monitoring, ECG, correlation dimension, Fractal analysis of temporary rows, non-linear dynamics of heart rhythm

Two levels spatial clustering

Arlette Antoni

Universite Europ\éenne de Bretagne, Universite de Bretagne-Sud, CNRS, Lab-STICC, Centre de Recherche Yves Coppens
BP 573, F-56017 Vannes cedex, France
Email:arlette.antoni@univ-ubs.fr

This paper deals with spatially constrained clustering. Clustering is spatially constrained if the cells composing a cluster and thus having similarity with regards to the measured variables also verify spatial homogeneity.
 In the literature, spatially constrained clustering is achieved using two types of methods :
- spatial constraints are applied after classical clustering,
- spatial constraints are applied within the clustering process.
 As explained in (Patil 2004) and (Urban 2004), the results of type I methods are more readily interpretable with respect to the variables, "but it tends to give small patches and splinter". On the other hand, type II methods yield tighter clusters, but are more problematic to interpret. The two methods often do not yield the same results.
 Here the objective is to optimize any classical clustering criterion while fixing as usual the number of classes and at the same time the number of regions. For this purpose, a two level stochastic bio-mimetic algorithm is provided and compared with other methods.
Keywords: Spatial clustering, Clustering, Spatial organization, Stochastic bio mimetic algorithms.

References

Patil G. P. and Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots, *Environmental and Ecological Statistics*, **11, 183-197**, 2004.

Urban D.L. Multivariate Analysis. Nonhierarchical Agglomeration. Spatially Constrained Classificatio. 2004.

http://www.env.duke.edu/landscape/classes/env358/mv_pooling.pdf

New technology in shopping: Forecasting electronic shopping with the use of a Neuro-Fuzzy System

GEORGE ATSALAKIS
Technical University of Crete, Department of Production Engineering and Management, Crete, Greece, atsalak@otenet.gr

 This paper presents the application of neuro-fuzzy techniques in forecasting a new technology in shopping. Neural networks have been used successfully to forecast time series due to the significant properties of treating non linear data with self learning capability. However, neural networks suffer from the difficulty of dealing with qualitative information and the "black box" syndrome that more or less limits their applications in practice. To overcome the drawbacks of neural networks, in this study we proposed a fuzzy neural network that is a class of adaptive networks and which is functionally equivalent to a fuzzy inference system. The results derived from the experiment based on the electronic sales indicate that the suggested fuzzy neural network could be an efficient system to forecast a new technology in shopping. Experimental results also show that the neuro-fuzzy approach outperforms the other two conventional models (AR and ARMA).
Keywords: new technology forecasting, electronic shopping forecasting, neuro-fuzzy forecasting, ANFIS

Spatio Temporal risk mapping based on hidden Markov random fields and variational approximations

Lamiae Azizi[1], David Abrial[2], Myriam Charras-Garrido[2], Florence Forbes[3]

[1]INRA/INRIA France
[2]INRA Clermont-Ferrand
[3]Mistis Team, INRIA Grenoble, France

Spatio Temporal analysis facilitates the detection of regional clusters exhibiting similar behavior, and identifies areas where patterns of risk are maintained over time.

The approach of Besag, York and Mollié (BYM) is a widely used model in spatial analysis, but it doesn't take into account the presence of discontinuities in the spatial structure of the risk and it ignores its temporal variation.

Following the approach based on finite mixture models proposed by Green and Richardson (2002), we propose to model the prior probabilities through a hidden Markov Random Field. Instead of using MCMC for inference, which would be very time consuming, we use the EM algorithm using Mean Field-Like approximations.

The goal of this work is to show the advantages of this approach using several simulation data and extend its use for the spatio-temporal modelling of epidemiological risk.

Réferences

Besag J., York J., Mollié A., 1991. Bayesian image restoration, with two applications in spatial statistics. *AISM, 43, 1-59*.

Green P., Richardson S., 2002. Hidden Markov models and disease mapping**.** *J AM STAT ASSOC. 97:1055-1070*.

Clustering Complex Heterogeneous Data Using a Probabilistic Approach

Helena Bacelar-Nicolau, Fernando Nicolau, Áurea Sousa and Leonor Bacelar-Nicolau

University of Lisbon, Faculty of Psychology, Lab. of Statistics and Data Analysis, Portugal, hbacelar@fpce.ul.pt
New University of Lisbon, FCT, Department of Mathematics, Portugal, clad@gmail.com
University of Azores, Department of Mathematics, Portugal, aurea@uac.pt
University of Lisbon, Faculty of Medicine, Lab. of Biomathematics & Institute of Preventive Medicine, Portugal, lnicolau@fm.ul.pt

In complex data bases, we are very often concerned with matrices where data units are described by a heterogeneous set of variables. Therefore the question arises of how we should measure the similarity between statistical data units in a coherent way, if different types of variables are involved. Traditionally partial similarity coefficients for each type of variables are computed, and then a convex linear combination of those similarities gives a global similarity between data units. Such procedure should be performed in a consistent way, combining comparable similarity coefficients in a valid / robust global similarity index. So far we have been using the so-called affinity coefficient (proposed by Matusita K. (1951), extended e.g. in H. Bacelar-Nicolau (1988) to cluster analysis framework) and generalized affinity coefficient for that purpose, respectively in two-way or in three–way / *symbolic* clustering contexts (e.g. in H.H. Bock and E. Diday (2000), pp160-165).

Here we apply the procedure to cases where heterogeneous variables of two pertinent types, are simultaneously present: interval and binary variables. We use a global probabilistic generalized affinity coefficient and compare a probabilistic approach (e.g. in I. C. Lerman (1981) and H. Bacelar-Nicolau (1988)) to the empirical one. Application on an example issued from the literature of symbolic data analysis illustrates how both coefficients work.

Keywords: similarity coefficient, interval variables, binary variables, hierarchical clustering model, probability distribution function, three-way / symbolic data.

Main References

1. Bacelar-Nicolau H., Two Probabilistic Models for Classification of Variables in Frequency Tables, *Classif. and Related Methods of Data Analysis*, H. H. Bock (ed.), North Holland, 181-186 (1988).
2. Bock H.H. and Diday E. [eds.], *Analysis of Symbolic Data Exploratory Methods for Extracting Statistical Information from Complex Data*. Springer (2000).
3. Lerman I. C., *Classification et Analyse Ordinale des Données*. Dunod (1981).
4. Matusita K., On the theory of statistical decision functions, *Ann. Instit. Stat. Math.*, *III*, 1-30 (1951).

A resampling based test for pairwise dependence in spatial processes

Bacro Jean-Noël[1], Bel, L[2] and Lantuéjoul, C[3]

1University of Montpellier II, [2]AgroParisTech, [3]MinesParisTech, France

The mean and the extremes of a spatial process are essentially of a different nature. Whereas the estimator of the mean tend to stabilize, the maximum keeps on jumping up as the support of the observations increases. As a consequence, standard geostatistic approaches are not appropriate to understand how extreme events tend to occur jointly or separately. Moreover, the maximum and the mean of a spatial process over a region may exhibit very different spatial dependencies. A concept of *asymptotic independence* has been introduced in multivariate extreme value theory. Roughly speaking, asymptotic independence means that each vector of componentwise maxima has its components ultimately independent. Fitting a model under a wrong dependence assumption in extreme values may result in a large extrapolation error. Here, we focus on the pairwise dependence modeling of maxima from spatial data. We propose to test asymptotic independence at any lag using a madogram obtained after transformation of the original field. In order to determine if the field is asymptotically independent at several lags, a multiple test procedure is derived, based on the bootstrap distribution of the number of times the null hypothesis is rejected. It is then tested on three classical spatial models and finally applied to two environmental datasets.

Time Series Segmentation by Cusum, AutoSLEX and AutoPARM methods

Ana Laura Badagián, Regina Kaiser and Daniel Peña

Department of Statistics, Universidad Carlos III de Madrid, España
abadagia@est-econ.uc3m.es; kaiser@est-econ.uc3m.es; dpena@est-econ.uc3m.es

Time series segmentation has many applications in several disciplines as neurology, cardiology, speech, geology and others. Many series in these fields do not behave as stationary and the usual transformations to linearity cannot be used. This paper describes and evaluates different methods for segmenting non-stationary time series.

We propose a modification of the algorithm in Lee et al. (2003) which is designed to searching for a unique change in the parameters of a time series process, in order to find more than one change using an iterative procedure. We evaluate the performance of three approaches for segmenting time series: AutoSLEX (Ombao et al., 2002), AutoPARM (Davis et al., 2006) and the iterative cusum method mentioned above and referred as ICM. The evaluation of each methodology consists of two steps. First, we compute how many times each procedure fails in segmenting stationary processes properly. Second, we analyze the effect of different change patterns by counting how many times the corresponding methodology correctly segments a piecewise stationary process.

ICM method has a better performance than AutoSLEX for piecewise stationary processes. AutoPARM presents a very satisfactory behavior. The performance of the three methods is illustrated with time series sets in neurology and speech.

Keywords: Time Series Segmentation, AutoSLEX, AutoPARM, Cusum Methods.

A new stochastic method to filter measurements on a random medium: Application to the fast estimation of atmospheric turbulent parameters

Christophe Baehr

Météo-France / CNRS and Toulouse Mathématique Institute, France christophe.baehr@meteo.fr

To denoise perturbed measurements made on a random medium, we have to use a stochastic representation of the measurement process. We have developed a new stochastic technique to model this process. We call our method the Acquisition Process of a random field along a random path. Using a local model of the medium with a conditioning to the sensor path, the Acquisition process gives a stochastic model available for stochastic filtering. Then we suggest to use a particle approximation to solve the associated filtering problem. Then we apply our specific particle filter to denoise corrupted atmospheric signal, more precisely real tridimensional atmospheric wind. The filter is able to retrieve the characteristic of the turbulent fluid. Then we compare classical large scale dissipation rate estimation with our fast estimation given by our filter. The next step of this work will be the integration of the mobile sensor platform dynamic.

Impulse Control of Partially Observed Diffusion

Fouzia Baghery-Kabbaj and Isabelle Turpin

Applied Mathematic Laboratory of Valenciennes (LAMAV), Valenciennes, France
fbaghery@univ-valenciennes.fr, iturpin@univ-valenciennes.fr

This paper concerns systems driven by partially observed diffusion process, and subjected to impulse control. Actually, the controller has only access to a part of the information of the signal's state (for example delayed observations occur, or only certain components of the state system can be observed). On the other hand, at random times θ_j (j=1,2...), he exerts jumps of intensity ξ_j over the system state. The impulse control is defined by means of the two sequences, θ_j and ξ_j which are his decision variables. We characterize the value function (which minimizes a given cost criteria), as well as the corresponding optimal policy.
To begin with we focus on mixed control (impulse and continue control) of a process, solution of Zakai-equation. Then, in presence of transaction cost, we define the value function by approximations.

Adaptive bandwidth selection in hazard rate estimation

D. Bagkavos and A. Kalamatianou

Abstract. We consider the issue of bandwidth selection in adaptive kernel based hazard rate estimation under the local linear framework when the data are randomly right censored. Specifically, the Akaike Information Criterion and the Empirical Bias Banwidths methods are extended from the density setting and evaluated numerically as potential bandwidth selectors when a kernel hazard rate estimate uses a variable, according to each data point, smoothing scheme. We conclude with graphical illustrations on distributional and real life datasets.
Keywords: Stochastic simulation, variable bandwidth, kernel, hazard rate, bandwidth selection, local linear.

Some Cure Rate Models and Associated Inference and Application to Cutaneous Melanoma Data

Narayanaswamy Balakrishnan

Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

In this talk, I will introduce some new cure rate models and discuss associated inferential procedures. I will illustrate the use of these models by applying them to a cutaneous melanoma data, and discuss the findings and their significance.

A fixed-interval smoothing algorithm for Jump-Markov State-Space Systems

Noémie Bardel and François Desbouvries

CITI Department and CNRS UMR 5157
TELECOM SudParis, 91011 Evry, France
noemie.bardel@it-sudparis.eu, francois.desbouvries@it-sudparis.eu

Let us consider the classical conditionally linear Gaussian model, also called Jump-Markov State-Space (JMSS) system, described by the equations:

$$\begin{cases} x_{n+1} = F_{n+1}(r_{n+1})x_n + v_n \\ y_n = H_n(r_n)x_n + w_n \end{cases},$$

in which $\{r_n\}$ denotes a discrete Markov Chain with known transition probabilities, and $\{v_n\}$, $\{w_n\}$, x_0 are Gaussian vectors which are independent, mutually independent and independent of $\{r_n\}$.

The Bayesian fixed-interval smoothing problem we address here consists in efficiently computing fixed-interval state estimates $E[x_n|y_{1:N}]$ and $E[r_n|y_{1:N}]$ with $n < N$. Let us start from the following factorization:

$$E[x_n|y_{1:N}] = \sum_{r_{n+1}} E[x_n|r_{n+1},y_{1:N}]\, p(r_{n+1}|y_{1:N}).$$

The first term in the sum can be approximated by a Kalman smoother, however the exact computation of the second term requires an exponential computational cost, so one needs to use suboptimal solutions. In Doucet *et al.* (*IEEE Tr. Signal Processing*, 2001), a particle smoother gives an approximation of $p(r_{n+1}|y_{1:N})$. However, to maintain diversity in the particles and avoid the problem of sample depletion, one has to include at each time a Monte Carlo Markov Chain (MCMC) step, which is computationally intensive.

In this paper we propose an alternative smoothing algorithm based on a different way to approximate $p(r_{n+1}|y_{1:N})$. More precisely, we assume that $(r_{1:N},y_{1:N})$ follows a Partially Pairwise Markov Chain (PPMC), whose joint distribution satisfies:

$$p(r_{1:N},y_{1:N}) = p(r_1,y_1)\prod_{n=1}^{N-1} p(r_{n+1},y_{n+1}|r_n,y_{1:n}).$$

The interests of this assumption are multiple. The model takes into account long memory observations; $p(r_{n+1}|y_{1:N})$ can be computed exactly by a forward-backward procedure; and the parameters of the PPMC which is closest to the true model $(r_{1:N},y_{1:N})$ can be estimated by an Iterative Conditional Estimation (ICE) method. To sum up, rather than using a Particle Smoother for approximating $p(r_{n+1}|y_{1:N})$, we approximate the stochastic model which describes the relation between $r_{1:N}$ and $y_{1:N}$, but we compute $p(r_{n+1}|y_{1:N})$ exactly.

TreeRank: a R package for bipartite ranking

Nicolas Baskiotis[1], Stéphan Clémençon[2], Marine Depecker[1,3], Nicolas Vayatis[1]

[1] CMLA, ENS Cachan, CNRS, UniverSud, Cachan France, nicolas.baskiotis@gmail.com
[2] Télécom ParisTech, Paris, France
[3] Renault DREAM-DTAA, Guyancourt, France

In the Bipartite Ranking setup, the goal is to learn a ranking function to sort data lying in a multidimensional feature space. Indeed, in a wide variety of fields, such as medical diagnosis, information retrieval or credit-risk screening, the goal of a binary supervised machine learning approach is less to predict the binary label of new data than to order them so that positive instances appear on top of the list with highest probability. TreeRank is a R package (Baskiotis et al. 2010) that implements the methodology introduced recently by Clémençon and Vayatis (2009). This algorithm uses a greedy recursive partitioning strategy in order to infer a piecewise approximation of the optimal scoring function. At each step, the partition of the remaining space is guided by a local optimization of the area under the ROC curve (AUC). The produced ranking function is summarized by a rooted, binary, left-right oriented tree.

  The presented R package includes the generic TreeRank algorithm as two version of this algorithm, one with a SVM inspired partitioning rule and the other one with a CART approach

(Breiman et al. 1984). The package includes as well tools for V-cross-validation, procedures for resampling and feature randomization, graphical user interfaces to setup and launch the algorithms, and graphical interactive tools to explore, manipulate and use the output ranking trees. Finally, the package includes an implementation of the algorithm proposed by Clémençon et al (2009) for solving multidimensional nonparametric two-sample problems. The TreeRank R package is available from now under GPL license on the official CRAN repository and at http://treerank.sourceforge.net/.
 Keywords: Bipartite ranking, ROC curve, R

References
1. Baskiotis, N., Clémençon, S., Depecker, M., Vayatis, N., R-implementation of the TreeRank algorithm. *Submitted for publication.*
2. Breiman, L., Firedman, J., Olshen, R., Stone, C., Classification and Regression Trees. Wadsworth and Brooks (1984).
3. Clémeçon, S., Depecker, M., Vayatis, N., AUC optimization and the two-sample problem, *Proceedings of the 2009 conference on Advances in neural information processing systems (NIPS 09)* (2009).
4. Clémeçon, S., Vayatis, N., Tree-based ranking methods. *IEEE Transactions on Information Theory,* **9**, 4316–4336 (Sept. 2009).

Decision approach for a Bayesian single variable sampling plan

Hacène Belbachir

University of Science and Tachnology Houari Boumediene, Algiers, Algeria
hbelbachir@usthb.dz

We consider a simple sampling plans model based on classical Bayesian decision theory with quadratic loss function. We introduce a decision function with random mixed region. We explicit the Bayes risk and suggest an algorithm for determining the optimal sampling plan.
Key Words: Bayesian decision theory, Statistical quality control, Single sampling plan.

Applying the Non-Proportional Odds Model to Set Multiple Cut Scores on a Test

Rosa Bersabé and Teresa Rivas

Faculty of Psychology, University of Málaga, Málaga, Spain, bersabe@uma.es, moya@uma.es

In this paper we use the non-proportional odds (NPO) model to set multiple cut-offs on a test score. The NPO model is an ordinal logistic regression model which is based on the relationship between a continuous test score (X) and an ordinal outcome variable (Y) with more than two categories. Cut scores are established at the test scores at the intersections of adjacent category distributions. The complete application of this analytical procedure is illustrated by an example with data from an actual study on eating disorders. In this example, two cut scores on the Eating Attitudes Test (EAT-26) are obtained in order to differentiate between three ordered categories: (1) asymptomatic; (2) symptomatic; and, (3) eating disorder. These categories resulted from the responses to a self-report (Q-EDD) that operationalises DSM-IV diagnostic criteria for eating disorders. Alternatives to the NPO model to set multiple cut-off scores are discussed.
Keywords: Standard-setting; cut-off scores; ordinal logistic regression; non-proportional odds model; psychometrics.

Financial Forecasting using Neural Networks

Dagmar Blatná, Jiří Trešl

University of Economics Prague, Czech Republic, blatna@vse.cz ; tresl@vse.cz

The possibility of application of neural networks for the prediction of both stock and exchange rate returns was investigated. First, the capability of neural networks to reveal specific underlying process was studied using different simulated time series. Second, actual weekly returns from Czech financial markets were analyzed and predicted. Particularly, the problems connected with capturing of outliers and structural breaks were discussed. The predictive power of neural networks was investigated both as a function of network architecture and the length of training set.
Key Words: Neural networks, financial time series, predictive power

Non-normality on the power of randomization tests: a simulation study using R

Fernando Branco[1] , Amilcar Oliveira[2,3] and Teresa Oliveira[2,3]

Universidade Lusófona, Lisboa[1]
DCeT[2] - Universidade Aberta, Lisboa
CEAUL[3]-Center of Statistics and Applications, University of Lisbon

The aim of the present research is to evaluate the impact of non-normality on the power of the randomization test for difference between the means of two independent groups. In literature, many studies such as we can see in Micceri, 1989; Mosteller & Tukey, 1968; Stigler, 1977; Wilcox, 1995a, 1995b; suggest that non-normality is common in research data sets. To manipulate non-normality, we used the set of 15 distributions used in Marron and Wand (1992)'s simulation study. These distributions, which have been used in quite a few subsequent studies, can all be written as mixtures of Gaussian distributions. They include distributions that are essentially normal with heavy tails and some outliers ('mild non-normality') and others with a more extreme non-normality, namely multi-modal distributions ('extreme non-normality'). We evaluated the power of the randomization test, and also the power of the Student-t test, as a comparison standard, with data simulated from the 15 Marron-Wand distributions for seven values of effect size and three sample sizes ($n_1 = n_2 = 8$, $n_1 = n_2 = 16$, $n_1 = n_2 = 32$). For each condition, we generated 20 000 samples, and for each one the power of randomization tests was estimated using 1 000 permutations. We set the value of Type I error probability at 0.05. The results show that, in terms of power, the two tests are similar, with a slight advantage for the randomization test over the Student-t test. When we compare the non-normal distributions with the Gaussian, we observe some gains in power in the case of 'mild non-normality' distributions and decreases in power in the case of 'extreme non-normality' distributions. These differences in power are inversely related with sample size.

Reformulation of general chance constrained problems using penalty functions

Martin Branda

Charles Univerzity in Prague/Univerzity of economics, Prague, Czech Republic
martin.branda@seznam.cz

We explore reformulation of nonlinear stochastic programs with several joint-chance constraints by stochastic programs with suitably chosen penalty-type objectives. We show that the two problems are asymptotically equivalent. Simpler case with one chance-constraint and particular penalty function was studied in Ermoliev et al. (2000). The

obtained problems with penalties and with a fixed set of feasible solutions are much simpler to solve and analyze than the chance constrained programs.
Key Words: stochastic programming, chance constrained problems, penalty function problems, asymptotic equivalence

A Simple and Powerful Test for the Stable Paretian Distribution

Simon Broda, Marc Paolella*

*Swiss Banking Institute, University of Zürich and Swiss Finance Institute (SFI), paolella@isb.uzh.ch

A new test is proposed for testing if a set of data is in the class of stable Paretian distributions.
The test is trivially computed, has correct size, and much higher power than an existing test. An application to financial returns data is presented.
Key Words: Empirical Finance, Stable Paretian

Conditioned random walks and statistical decision for tests with large p-values

Michel Broniatowski

LSTA Université Paris 6, Paris, France, michel.broniatowski@upmc.fr

This talk presents a sharp approximation of the density of long runs a random walk conditioned on its end value as the number of summands tends to infinity. The conditioning event is a moderate deviation event. The result extends the Gibbs conditional result in the sense that it provides a description of the distribution of the random walk on long subsequences. A simulation algorithm for these random walks is presented. As a by product a new methodology aiming at assessing decisions when a test statistics has a p- value of small order under a given simple null hypothesis is proposed. It is based on the comparison of the observed dataset with typical datasets generated under the null hypothesis conditioned on the same value of the statistics as calculated on the observed dataset.
Key Words: Conditioned random walk; test; moderate deviation; simulation

Ranking teams in the NCAA Division 1 Men's Basketball Tournament.

Mark Brown

Department of Mathematics, City College, CUNY

The NCAA division 1 men's basketball tournament is one of the most popular sporting events in the USA. There are several methods of comparison of teams that are available to fans and bettors. These include the AP poll of sportswriters, the ESPN/USA Today coach's poll. The RPI (ratings percentage index), the Sagarin ratings, the tournament seedings, the Las Vegas pointspreads, and one which I am involved with, the LMRC model. The LMRC (logistics regression Markov chain) model, introduced by Professors Paul Kvam and Joel Sokol (NLRQ, 2006) has been very successful in statistical tests against the above methodologies and has received a good deal of media attention. I'll talk about an updated version, which incorporates empirical Bayes methodology. It was developed by the speaker and Joel Sokol of Georgia Tech University, in a paper that will appear in the Journal of Quantitative Analysis in Sports.

Forecasting the Incidence of Cancer in Regional Victoria, Australia

Nicholas Brown, Philippa Hartney, Rachael Hamilton-Keene and Terry Mills

Loddon Mallee Integrated Cancer Service, Bendigo, Victoria, Australia,
tmills@bendigohealth.org.au

Loddon Mallee Integrated Cancer Service (LMICS) is responsible for planning the delivery of cancer services in the Loddon Mallee Region of Victoria, Australia. Forecasting the incidence of cancer in the region plays a key role in strategic planning for these services. In this paper, we describe the context of our work, present a review of the literature on forecasting the incidence of cancer, discuss contemporary approaches to the problem especially functional data analysis, describe our experience with the models at LMICS, and list special issues associated with applying these models in regional Australia. The extensive bibliography illustrates the world-wide interest in this forecasting problem.
Keywords: Stochastic models, Functional data analysis, Poisson regression, Strategic planning, Health care

New approach to dynamic XL reinsurance

Ekaterina Bulinskaya and Aleksander Gromov

Department of Mathematics and Mechanics, Moscow State University, Moscow, Russia
ebulinsk@mech.math.msu.su, gromovaleksandr@gmail.com

We consider an insurance company having a possibility to choose and buy dynamically an unlimited excess of loss reinsurance. We model risk process R_t of the company by a Cramer-Lundberg process with claim arrival intensity λ and absolutely continuous claim size distribution F. Retention level b and limiting level (b+M) are assumed to be chosen dynamically, i.e. we consider predictable strategies $Z_t=(b_t,M_t)$ and the insurer can adjust retention b_t and limit b_t+M_t, at any time $t \geq 0$ on the base of information available before t.

Let T_i be the occurrence time of the i-th claim, W_i the claim amount, c the premium intensity of the insurer, N_t the number of claims in time interval (0,t], and assume that the reinsurer uses the expected value principle with safety loading $\theta > 0$, then the risk process under the strategy $Z_t=(b_t,M_t)$ is

$$R_t^Z = s + ct - (1+\theta)\lambda \int_0^t E\min\{M_x, \max(0, W-b_x)\}dx -$$

$$- \sum_{i=1}^{N_t} (\min\{W_i, b_{T_i}\} + \max\{0, W_i - b_{T_i} - M_{T_i}\})$$

where s is the initial surplus.

Our aim is to maximize the survival probability of the insurer. To this end we calculate the function $\delta(s)=\sup_Z\{\delta_Z(s)\}$, where $\delta_Z(s)$ is the survival probability of the cedent, i.e. $\delta_Z(s)=P(R_t^Z \geq 0, t \geq 0 | R_0^Z=s)$. The corresponding Hamilton-Jacobi-Bellman equation for our optimization problem is

$$\delta'(s) = \inf_{b>0, M>0} \lambda \frac{\delta(s) - E[\delta(s - \min\{W,b\} - \max\{0, W-b-M\})]}{c - (1+\theta)\lambda E[\min\{M, \max(0, W-b)\}]}$$

Theorem. *There exists a solution δ(s) of this equation, continuous on [0,∞) and continuously differentiable on (0,∞). Moreover, the strategy based on $Z^*(s)=(b^*(s),M^*(s))$ providing min in the right-hand side of the above equation for δ(s) maximizes the survival probability.*

This theorem states the existence of the optimal strategy in case when insurer has a possibility to choose and buy XL contract at any time. Also in this paper, we consider the case when the insurer can buy XL contract only at the end of each year. This leads to another control problem and existence theorem.

To support the theory we provide some useful numerical examples.

Discrete time models with dividends and reinsurance

Ekaterina Bulinskaya and Daria Yartseva

Department of Mathematics and Mechanics, Moscow State University, Russia
ebulinsk@mech.math.msu.su, yartseva@gmail.com

We consider a discrete time model of insurance company capital. It is supposed that company uses the barrier strategy of dividends payment and quota share reinsurance. The claims have continuous distribution with density function $p_\xi(x)$. It is established that average costs satisfy the equation $g(x) = Ag(x)$. The mapping $A : C[0,b] \to C[0,b]$ is given by

$$Ag(x) = \beta g(b) \int_0^{f_1(x)} p_\xi(y)dy + \beta \int_{f_1(x)}^{f_2(x)} g(h(x,y))p_\xi(y)dy + f_3(x) \quad (1)$$

where $0 < \beta < 1$, b is the dividend barrier and $f_1(x), f_2(x), f_3(x)$ are some special continuous functions. It could be shown that A is a contraction mapping.

Theorem 1 Let A be a contraction mapping of the following form

$$Ag(x) = f(x)g(b) + \int_0^b K(x,y)g(y)dy$$

where $K(x,y) \geq 0$ and $f(x)$ are continuous functions. If function $\varphi(x)$ satisfies the inequality $\varphi(x) \geq A\varphi(x)$ then $\varphi(x) \geq 0$.

Theorem 2 Let A be the contraction mapping from Theorem 1 and $m(x)$ the solution of equation $m(x) = Am(x) + h(x)$. If $\varphi(x)$ satisfies the inequality $\varphi(x) \geq A\varphi(x) + h(x)$ then $\varphi(x) \geq m(x)$.

Equation for average time until ruin is a particular case of (1) with $\beta = 1$ and $f_3(x) = 1$. The properties of solutions of these equations are studied. The method of finding upper and lower bounds for solutions is proposed.
In some special cases the reinsurance strategy that maximizes average time until ruin is obtained and exact solutions are provided.
Keywords: Upper and Lower Bounds, Continuous Claims Distribution, Optimal Reinsurance

On Imbedding Quantum Channels in Time-Continuous Markov Quantum Processes

Andrey Bulinski

Department of Mathematics, Moscow Institute of Physics and Technologies,
Moscow, Russia, bulinski@mail.mipt.ru

The classical counterpart of the imbedding problem was studied in the context of finite state Markov chains. S.Johansen established criteria for imbedding a finite stochastic matrix in a homogeneous and non-homogeneous time-continuous Markov chain. They involve the notion of infinite divisibility and other related concepts such as infinite factorizability for these matrices.

In the context of quantum dynamics and quantum information theory, when dealing with systems having finite-dimensional underlying Hilbert space H, one considers completely positive (CP) stochastic dynamical maps or channels. They correspond to maps of states (densities), i.e. preserve trace. This Schroedinger description is accompanied by the dual, Heisenberg picture, treatment of unital CP maps of the algebra $B(H)$ isomorphic to $M_n(C)$. The Markov time-continuous dynamics corresponds to a quantum dynamical semi-group of such maps.

A.S.Holevo introduced a concept of infinite divisibility for quantum channels. M.M.Wolf and J.I.Cirac proposed criteria for imbedding a finite-dimensional CP into a Markov homogeneous evolution with time-independent as well as time-dependent generator. They introduced the concepts of infinitesimal divisibility and other related notions for CP maps over $M_n(C)$.

The goal of our study is to extend these results to infinite-dimensional case and more generally to mappings of arbitrary von Neumann algebras and their preduals.

Keywords: Quantum Channel, Markov Quantum Process

References
1. Bulinski, A. Noncommutative dynamical systems, quantum stochastic processes and modeling of quantum channels, in *Proc. of the 6th St. Petersburg Workshop on Simulation* (S.M.Ermakov et al. eds), St. Petersburg State Univ., 83-87, (2009).
2. Johansen, S. A CLT for finite semigroups and its application to the imbedding problem for finite-state Markov chain. *Z. Wahrsch. verw. Geb*, **26**, 171-190 (1973).
3. Wolf, M.M. and Cirac, J.I. Dividing quantum channels. *Commun. Math. Phys*., **279,** 147-169 (2008).
4. Størmer, E. Separable states and positive maps. *J. Funct. Anal*. , **254**, 2303-2312 (2008).

Some Results on the Fuzzy Least Square Regression Model with fuzzy intercept

Francesco Campobasso and Annarita Fanizzi

Department of Statistical Sciences "Carlo Cecchi", University of Bari, Italy,
fracampo@dss.uniba.it, a.fanizzi@dss.uniba.it,

Fuzzy regression techniques can be used to fit fuzzy data into a regression model, where the deviations between the dependent variable and the model are connected with the uncertain nature either of the variables or of their coefficients. P.M. Diamond (1988) treated the case of a simple fuzzy regression of an uncertain dependent variable on a single uncertain independent variable, introducing a metrics into the space of triangular fuzzy numbers. Recently we provided some theoretical results (5) about the decomposition of the sum of squares of the dependent variable of the simple regression model obtaining the expression of two additional components of variability, besides the regression deviance and the residual one, which arise from the inequality between theoretical and empirical values of the average fuzzy dependent variable.
In this work we demonstrate how the estimation procedure of a simple fuzzy regression model changes depending on whether the intercept is fuzzy or not, also with reference to the the components of variability that characterize such a model. In particular we explicit the expression of the estimated intercept - including the center and the left and right spreads – and we verify that the sum of squares of the dependent variable consists only in the sum of the regression and the residual deviances, like in the classic OLS estimation procedure.

In order to appreciate the obtained results, we propose a comparison between different models in an application case.

Keywords: Fuzzy least square regression, multivariate generalization, total deviance, decomposition, goodness of fitting.

References
1. Campobasso, F., Fanizzi, A., Tarantini, M., Fuzzy Least Square Regression, *Annals of Department of Statistical Sciences, University of Bari, Italy*, 229-243 (2008).
2. Diamond, P. M., Fuzzy Least Square, *Information Sciences*, 46:141-157 (1988).
3. Kao, C., Chyu, C.L., Least-squares estimates in fuzzy regression analysis, *European Journal of Operational Research*, 148:426-435 (2003).
4. Takemura, K., Fuzzy least squares regression analysis for social judgment study, *Journal of Advanced Intelligent Computing and Intelligent Informatics*, 9(5), 461:466 (2005).
5. Campobasso, F., Fanizzi, A., Tarantini, M., Some results on a multivariate generalization of the Fuzzy Least Square Regression, *Proceedings of the International Conference on Fuzzy Computational* (2009).

Complex modelling and hierarchical control of hydraulic systems

Enrico Canuto[1], Wilber Acuna-Bravo[1] and Manolis Christodoulou[2]

[1] Politecnico di Torino, Dipartimento di Automatica e Informatica, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, enrico.canuto@polito.it, wilber.acunabravo@polito.it
[2] Technical University of Crete, Chania, Crete, Greece, 73100 and Politecnico di Torino, Dipartimento di Automatica e Informatica, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, manolis@polito.it

The paper outlines the general control problem of a complex hydraulic circuit, either fixed or mobile, where a variable flow pump must meet the demand of a finite number of hydraulic variable loads. They may comprise hydraulic cylinders (motors) moving a load along a reference trajectory, primary lubricating (steady) flows, secondary flows. The pump flow is distributed to motor volumes through solenoid-driven proportional valves. Cylinder reference flow must meet motor load reference rate, whereas load resistance and inertial forces impose cylinder pressure. Proportional valves ensure, within their limits, flow to cylinders by regulating valve aperture, thus performing a low-level (local) control task. Higher level task must ensure a sufficient pressure drop from pump output volume and supply line to cylinder volumes, in presence of variable cylinder pressure range and rate. Pressure rates may be rather high because of sudden load resistance. Moreover, cylinders and secondary flows must be ranked within the max available pump flow. Pressure range is limited by relief valves, not treated here.

First essential dynamics of the main hydraulic elements common to cylinders and variable flow pumps are recalled, and simplified through smooth singular perturbation in view of the control embedded model. Then hierarchical control problem is formulated and solved within the Embedded Model Control architecture. Simulated results are shown for the higher-level task.

Keywords: Hydraulics, hierarchical control, complex modeling, variable flow pumps, proportional valves

Handling unreliable follow-up in lifetime data analysis

C. Caroni

Department of Mathematics, School of Applied Mathematical and Physical Sciences,
National Technical University of Athens, Athens, Greece, ccar@math.ntua.gr

This paper examines several topics whose common theme is the effect of unreliable follow-up information in a study that examines times to an event or between recurrent events. One problem may be undetected withdrawal of the unit from the study. This results in incorrectly recording the interval from the last recorded event until the end of the study as a right-censored event time. A test is developed for excessively large values of these times, using approximations to the sum of Beta random variables. This test does not handle units with no events; instead a randomization test is developed for this case. If these right-censored intervals until the end of study are thought to be unreliable, each unit's data can be treated instead as ending at the last recorded failure. The implications of treating data in this way are examined. Finally, follow-up data may be seriously unreliable if the events are recorded but not linked to the correct units. The effects of this distortion of data are investigated.

Optimal Pricing of a Conspicuous Product During a Recession that Freezes Capital Markets

Jonathan P. Caulkins[1], Gustav Feichtinger[2], Dieter Grass[3], Richard F. Hartl[4], Peter M. Kort[5] and Andrea Seidl[6]

[1]Carnegie Mellon University, Pittsburgh, USA, caulkins@cmu.edu
[2]Vienna University of Technology, Vienna, Austria, gustav@eos.tuwien.ac.at
[3]The Beijer Institute of Ecological Economics, Stockholm, Sweden, dieter@eos.tuwien.ac.at
[4]University of Vienna, Vienna, Austria, richard.hartl@univie.ac.at
[4]Tilburg University, Tilburg, Netherlands, kort@uvt.nl
[5]Vienna University of Technology, Vienna, Austria, aseidl@eos.tuwien.ac.at

The paper considers the problem of how to price a conspicuous product, when the economy is in recession that disrupts capital markets. A conspicuous product in this context is a luxury good for which demand is increasing in brand image. A high brand image implies that displaying consumption of the good to observers adds to the reputation of the consumer. Typically, brand image is built up when the good is priced to make it exclusive.
The recession is modeled here as having two effects: it reduces demand and it freezes capital markets so borrowing is not possible. In pricing the conspicuous product the firm faces the following trade-off. On the one hand it feels pressure to reduce the price to maintain sales volume in the face of reduced demand. On the other hand, reducing the price damages brand image and thus long term demand.
The paper analyzes the firm's pricing policy facing scenarios of mild, intermediate and severe recessions, while taking the threat of bankruptcy into account. It will be shown that a decision maker has to adapt the pricing strategy according to the severity of the recession and the initial brand image. For an intermediate recession the optimal solution is history-dependent.
Keywords: optimal control, multi-stage modeling, stochastic recession

Inquiry System for the School Bus Position and Arrival Time Information

Huey-Kuo Chen and Che-Jung Wu

National Central University, Taoyuan County, Taiwan, ncutone@ncu.edu.tw
National Central University, Taoyuan County, Taiwan, aoutow@gmail.com

School bus is a convenient and efficient transportation mode. Useful travel time information can help commuters make good choice of departure time as well as other travel decisions.

However, providing real time travel time information for an en route school bus is a difficult task as travel time information is uncertain due to external interferences. Hence no consistent tendency can be easily observed. In this research, we develop a novel approach that combines grey theory (for travel time prediction) and empirical decomposition method (for data analysis) to predict the arrival time at each stop along the school bus route. The experiments showed that our prediction approach employing real time data collected from geographic positioning system as well as historic data outperforms the approach involving historic data only. In addition, sensitivity analysis is performed for different number of stops used in the prediction process.
Keywords: Travel Time, Hilbert-Huang Transform, Empirical Mode Decomposition, Grey Theory.

The Spike Noise Based on the Renewal Point Process and Its Possible Applications

Olga A. Chichigina, Alexander A. Dubkov, Bernardo Spagnolo, and Davide Valenti

Physics Department, Lomonosov Moscow State University, Moscow, Russia, chichigina@hotmail.com
Radiophysics Department, N.I. Lobachevsky State University, Nizhniy Novgorod, Russia, dubkov@rf.unn.ru
INFM and Dipartimento di Fisica e Tecnologie Relative, Group of Interdisciplinary Physics Universita di Palermo, Palermo, Italy, spagnolo@unipa.it, valentid@gip.dft.unipa.it

We consider a non-Markovian random process in the form of spikes train, where the time intervals between neighboring delta-pulses are mutually independent and identically distributed, i.e. represent the renewal process (1967). This noise can be interpreted as the derivative of well-known continuous time random walk (CTRW) model process with fixed value of jumps. The closed set of equations for the characteristic functional of the noise, useful to split the correlations between stochastic functionals (2008), is obtained. In the particular case of Poisson statistics these equations can be exactly solved and the expression for the characteristic functional coincides with the result for shot noise (2005). Further we analyze the stability of some first-order system with the multiplicative spike noise. We find the momentum stability condition for arbitrary probability distribution of intervals between pulses. The general condition of stability is analyzed for the special probability distribution of intervals between pulses corresponding to so-called dead-time-distorted Poisson process. It means that within some time interval after each delta pulse the occurrence of new one is forbidden (like in neurons). The possible applications of the model to some problems of neural dynamics, epidemiology, ecology, and population dynamics are discussed.

References
1. Cox, D. R., Renewal Theory, New York, Chapman and Hall (1967).
2. Dubkov, A. A., Spagnolo, B., and Uchaikin, V. V., Levy flight superdiffusion: An introduction. *Int. J. Bifur. Chaos*, **18**, 2649-2672 (2008).
3. Dubkov, A. A. and Spagnolo, B., Generalized Wiener process and Kolmogorov's equation for diffusion induced by non-Gaussian noise source. *Fluct. Noise Lett.,* **5**, L267-L274 (2005).

Application of the computer simulation technique for investigating problems of parametric AFT-model construction

Ekaterina V. Chimitova and Natalia S. Galanova

Department of Applied Mathematics, Novosibirsk State Technical University, Novosibirsk, Russia, chim@mail.ru, Natalia-Galanova@yandex.ru

This paper is devoted to the investigation of the problems of selecting distribution law for time-to-event data obtained by accelerated life testing. The problems of parametric model verification based on testing goodness-of-fit to a specified distribution law by samples of

residuals have been considered. By means of computer simulation methods the statistical properties of model parameters estimators have been investigated depending on sample size, censoring degree and plan of experiment. The statistic distributions and the power of nonparametric goodness-of-fit tests have been studied. The Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises-Smirnov goodness-of-fit tests have been considered.

Keywords: Monte-Carlo method; parametric AFT-model; censored samples; maximum likelihood estimation; the Kolmogorov-Smirnov, the Anderson-Darling, the Cramer-von Mises-Smirnov goodness-on-fit tests

Dynamic Factor Analysis Models with Time-Varying Parameters

Sy-Miin Chow, Jiyun Zu and Guangjian Zhang

Dept. of Psychology, University of North Carolina, Chapel Hill, NC, USA,
symiin@email.unc.edu

Dynamic factor analysis models are one way of representing change processes through the combination of a time series model of choice and a factor analytic model. We consider dynamic factor analysis models with time-varying parameters in an application involving human emotion processes. Building on a dynamic factor analysis model with a vector autoregressive process at the factor level, we allow the corresponding cross-regression parameters to vary over time to evaluate whether and in what ways the linkages between human emotion factors change over time. Doing so requires some of the time series parameters to be represented as latent variables, which introduces nonlinearity into the model. We propose using the extended Kalman smoother (EKS) to estimate factor scores and track time-varying parameters and the Gaussian maximum likelihood (GML) for estimating other time-invariant parameters. Due to the finite length! of the data, standard asymptotic theory used to derive standard errors may not be tenable. A bootstrap procedure is used to estimate the standard errors and other distributional properties of the time-invariant parameters. Simulated and empirical examples are used to illustrate the performance of the proposed method and the bootstrap standard errors in comparison to asymptotic standard errors. Practical implications of fitting dynamic models with time-varying parameters to multivariate time series data of finite lengths are discussed.

Key Words: Bootstrap, nonlinear dynamic models, discrete-time stochastic processes

Two optimality results about sample paths properties of Operator Scaling Gaussian Random Fields

M. Clausel and B. Vedel

Paris 12 University, Créteil, France, clausel@univ-paris12.fr

Random fields are now used for modeling in a wide range of scientific areas including physics, engineering, hydrology, biology, economics and finance. An important requirement is that the data thus modelled present strong anisotropies which therefore have to be present in the model. Many anisotropic random fields have therefore been proposed as natural models in various areas such as image processing, hydrology, geostatistics and spatial statistics. In many cases, Gaussian models have turned to be relevant when investigating anisotropic problems. More generally anisotropic Gaussian random fields are involved in many others concrete situations and then arise naturally in natural model.

In many situations, the data present invariant features across the scales and several extensions of self-similarity property in an anisotropic setting have been proposed. Here we are interested in an anisotropic self-similar class of Gausssian fields: operator scaling Gaussian fields. Having in mind the problem of the estimation of the parameter, we study the sample paths properties of Operator scaling Gaussian random fields. Some characteristic properties of the anisotropy are revealed by the regularity of the sample paths. The sharpest way of measuring smoothness is related to these anisotropies and thus to the geometry of these fields

Key Words: Operator scaling Gaussian random field, anisotropy, sample paths properties

Analyzing time-course microarray data using functional data analysis - a review.

Norma Coffey[1] and John Hinde[1]

[1]Science Foundation Ireland funded BIO-SI Project,
School of Mathematics, Statistics & Applied Mathematics, National University of Ireland,
Galway, Ireland.

Gene expression data has typically consisted of a number of measurements made on many genes at a small number of points in time. These data tended to have lots of missing values, were measured possibly with a large amount of error, etc. Many multivariate techniques such as principal components analysis, mixed effects modelling and cluster analysis methods were used to model such data. These methods tended not to account for the correlation between measurements made on the same gene, facilitate the removal of noise from the measured data or account for any smoothness that may be evident in the expression profiles.

As data collection procedures have improved, gene expression profiles contain much larger numbers of time points than previously experienced. As a result, gene expression over time can be viewed as a continuous process and therefore represented as a continuous curve or function. Functional data analysis (FDA) is a statistical methodology developed to analyze functional data. Treating the data as functions rather than discrete values allows for the inclusion of non-uniformly sampled data, enables the experimenter to estimate expression values at times different from those used in the original experiment, impute missing values and remove noise from the measured data. Another key advantage FDA is the availability of derivative information. The use of derivatives (and dynamical systems theory) sets FDA apart from other multivariate techniques and extends the power of FDA over such methods. Since gene-expression can be thought of as part of a biological system, it could be suggested that use of the derivative may provide some additional insight into the behaviour of gene-expression data.

Many multivariate techniques, e.g. principal components analysis, regression analysis, etc. have been extended to the functional case. There have been several applications of FDA methodologies to gene expression data in recent years. These include functional principal components analysis, functional discriminant analysis and functional regression. These have been used to satisfy some of the main aims in modelling gene expression data, e.g. dimension reduction and clustering to determine groups of co-expressed genes, tests for differential expression between genes across treatment groups, discrimination and classification of genes, etc. This paper presents a review of the applications of FDA techniques to time-course gene expression data.

Strategies for Energy Management in Home Automation Systems

G. Conte, A. Perdon, D. Scaradozzi, M. Rosettani, G. Morganti, F. Dezi

Dipartimento di Ingegneria Informatica, Gestionale e dell'Automazione, Università Politecnica delle Marche – Ancona, Italy
gconte@univpm.it, perdon@univpm.it, d.scaradozzi@univpm.it,
morganti@diiga.univpm.it, rosettani@diiga.univpm.it.

This paper deals with the problem of resource management and allocation in home automation systems. In many situations, the global, concurrent demand of resources, like electricity and gas, by home appliances may exceed actual availability and this originate conflicts, whose solution requires the use of suitable control strategies. Here, we summarize and discuss different control strategies, which have been proposed by the authors in recent papers. Performances are compared with respect to generic scenarios and advantages and drawbacks of each one are briefly discussed.
Keywords: Home automation, Resource management, Optimization

A Mann-Whitney spatial scan statistic for continuous data

Cucala Lionel

Institut de Mathématiques et Modélisation, Université des Sciences et Techniques du Languedoc Montpellier, France

A new scan statistic is proposed for identifying clusters of high or low values in georeferenced continuous data. On the one hand, it relies on a concentration index which is based on the Mann-Whitney statistic and thus is completely distribution-free. On the other hand, the possible spatial clusters are given by an original graph-based method.
This spatial scan test seems to be very powerful against any arbitrarily-distributed cluster alternative.
These results have applications in various fields, such as the epidemiological study of rare diseases or the analysis of astrophysical data.

Some Dimension Reduction Methods in non-parametric spatial modelling

Sophie Dabo-Niang[1] Jean-Michel Loubes[2] and Anne-Françoise Yao[3]

[1] Laboratoire EQUIPPE, University Lille 3, Domaine du pont de bois, BP 60149, 59653 Villeneuve d'ascq cedex, France, sophie.dabo@univ-lille3.fr
[2] University Toulouse 3, Institut de Mathématiques de Toulouse, Equipe de Statistique et Probabilités, route de Narbonne, 31062 Toulouse, loubes@math.univ-toulouse.fr
[3] Laboratoire LMGEM, University Aix-Marseille 2, Campus de Luminy, case 901, 13288Marseille cedex 09, anne-francoise.yao@univmed.fr

In many areas: geology, oceanography, econometrics, soil science, epidemiology, physics, environment, risk management, image processing, the data are spatially dependent and their treatment require specific tools provided by spatial statistic. This work deals with the spatial regression. Unlike the parametric case, nonparametric spatial regression is only tackled in a few papers, among them Biau and Cadre[1], Dabo-Niang and Yao[2] , Halin et al.[3]. Their results show that, as in the *i.i.d.* case, the spatial nonparametric estimator of the regression function is penalized by the dimension of the regressor. More precisely, given observations of a random field $Z_i = (X_i;\ Y_i)$, i $\in$ N^N ($N \geq 1$), we investigate estimation of the regression function $Y_i = m(X_i)$ when the Y_i's are a real-valued and the X_i's are valued in a separable space (ε). Our estimation methods are based on the dimension reduction assumption:

$$m(x) = g(\Phi, x) \qquad (1)$$

where Φ a linear orthogonal mapping and g an unknown function defined from R^D to R, with D as small as possible. The function $g(.)$ being estimated by using a spatial kernel method. Some asymptotic properties (with rates) of our techniques will be checked and simulations with R will illustrate our results.

Keywords: Random fields, Spatial statistic, Functional data analysis, Dimension reduction.

References
1. Biau, G. and Cadre, B. (2004). Nonparametric Spatial Prediction, Statistical inference for Stochastic Processes, Vol. 7, pp. 327-349.
2. Dabo-Niang, S. and Yao, A.F. (2007). Kernel regression estimation for continuous spatial processes, Mathematical Methods of Statistics., 16, pp. 298-317.
3. Hallin, M., Lu, and Tran, L.T. (2004). Local linear spatial regression, Annal of Statistics, Vol. 32, No. 6, pp. 2469-2500.

Robust quantile estimation and prediction for spatial processes

Dabo-Niang, S. and Thiam, B.

Laboratoire EQUIPPE, Université Charles De Gaulle, Lille 3, maison de la recherche, domaine du pont de bois, BP 60149, 59653 Villeneuve d'ascq cedex, France.
sophie.dabo@univ-lille3.fr, baba.thiam@univ-lille3.fr

Spatial quantile estimation is an interesting and crucial problem in statistical inference for a number of applications where the influence of a vector of covariates on some response variable is to be studied in a context of spatial dependence. We present a statistical framework for modeling conditional quantiles of spatial processes assumed to be strongly mixing in space. We are mainly concerned in this work with L1 consistency as well as asymptotic normality of the kernel conditional quantile estimator in the case of random fields. We propose a spatial nonparametric predictor based on the conditional quantile estimator. We illustrate the proposed methodology with some simulations.
Keywords: Spatial processes, Kernel estimate, Conditional quantile, Spatial prediction.

References
1. Biau, G. and Cadre, B. (2004). Nonparametric spatial regression, Statistical Inference for Stochastic Processes, vol. 7, pp. 327-349.
2. Carbon, M., Tran, L.T. and Wu, B. (1997). Kernel density estimation for random fields, Statistics and Probability Letters, vol. 36, pp. 115-125.
3. Collomb, G. (1980). estimation non paramétrique de probabilités conditionnelles, C. R. Acad. Sci. Paris Sér I Math., vol. 291, pp. 427-430.
4. Dabo-Niang, S. and Yao, A. F. (2007). Spatial kernel regression estimation, MAthematical Methods of Statistics, vol. 16, Number 4, pp. 1-20.
5. Gannoun, A., Saracco, J. and Yu, K. (2003). Nonparametric prediction by conditional median and quantiles, Journal of Statistical Planning and Inference, vol 117, pp. 207-223.
6. Laksaci, A. Lemdani, M. and Ould-Said E. (2009). A generalized L1 approach for a kernel estimator of conditional quantile with functionnal regressors: Consistency and asymptotic normality. Statistics and Probability Letters, vol 79, Issue 8, pp. 1065-1073.
7. Matzner-Løber (1997). Prévison nonparamétrique des processus stochastiques. Thèse Université montpellier II.
8. Roussas, G.G. (1969). Nonparametric estimation of the transition distribution function of a Markov process, Ann. Math. Statist., vol. 40, pp. 1386-1400.
9. Tran, L. T. (1990). Kernel density estimation on random fields, Journal of Multivariate Analysis, vol. 34, pp. 37-53.
10. Yu, K. and Jone, M.C. (1998). Local linear quantile regression. J. Amer. Statist. Assoc. vol. 93, pp. 228-237.

Semi-Markov Disability Insurance Models

Guglielmo D'Amico, Montserrat Guillen and Raimondo Manca

University "G. D'Annunzio", via dei Vestini 31, Chieti 66013, ITALY, g.damico@unich.it

In this paper, we present a stochastic model for disability insurance contracts. The model is based on a discrete time non-homogeneous semi-Markov process to which the backward recurrence time process is joined. This permits to study in a more complete way the disability evolution and to face in a more effective way the duration problem. The use of semi-Markov reward processes gives the possibility of deriving equations of the prospective reserves and to investigate the complete-risk loss function and the reduced-risk loss function. The model is applied to a sample of contracts drawn at random from a mutual insurance company.
Key Words: Stochastic modeling, reward process, semi-Markov chain, loss function

SOME REMARKS ON THE HIPPARCOS SATELLITE PROPER MOTIONS DATA

Goran Damljanovic

Astronomical Observatory, Volgina 7, 11060 Belgrade 38, Serbia,
gdamljanovic@aob.bg.ac.rs

The Hipparcos Catalogue (as a result of the satellite HIPPARCOS ESA mission) with astronomical data for near 120000 stars was published in 1997. After that, a few new astrometric catalogues appeared (TYCHO-2, ARIHIP, etc.) because the investigations of the accuracy and systematic errors of the Hipparcos data noticed some unresolved astrometrical problems. Some of new catalogues are a good combination of the Hipparcos and ground-based data; like that it is possible to get more accurate coordinates and proper motions of stars than the Hipparcos Catalogue ones. Also, the new Hipparcos reduction was done by van Leeuwen, and the re-reduced Hipparcos data appeared in 2008.

The EOC-2 catalogue was finished (during last few years) with the results about improvement of Hipparcos coordinates and proper motions by using the astrometric observations of latitude/universal time variations. It was done via observed stars referred to Hipparcos Catalogue, together with the Hipparcos data. Mentioned ground-based data were used to determine the Earth Orientation Parameters (EOP) during the last century (there are a lot of optical observations made worldwide at many observatories). Now, they are useful to improve the accuracy of coordinates and proper motions of Hipparcos stars. In accordance with it, I used the latitude data of 26 instruments and combined them with the Hipparcos Catalogue ones (with suitable weights) to get the proper motions in declination for 2347 stars. I did it in order to improve the proper motions in declination for these Hipparcos stars. The results are compared with the other catalogues (the EOC-2, the Hipparcos, the re-reduced Hipparcos, etc.) ones via the values of standard errors, systematic errors, etc. My results are in good agreement with the other ones, but it is necessary to get about 20 years and more than 20 years of mentioned ground-based astronomical observations. The Least Squares Method with the linear model was used. Some results are presented here.

Predicting Caloric Efficiency in Turkeys Using the Group Method of Data Handling-Type Neural Networks

H. DARMANI KUHI[1]*, A. FARIDI[1], M. MOTTAGHITALAB[1], J. FRANCE[2]

[1]Department of Animal Science, Faculty of Agricultural Science, University of Guilan, PO Box 41635-1314, Rasht, Iran; * Darmani_22000@yahoo.com
[2]Centre for Nutrition Modelling Department of Animal & Poultry Science University of Guelph, Guelph, Canada

There has been extensive research conducted to clarify protein, essential AA and energy requirements in turkeys. Conventional laboratory and field-based techniques for determining nutrient requirements are expensive, cumbersome and time-consuming. These disadvantages have prompted a search for alternative methods. In determining nutrient requirements, the potential benefits from modeling growth in turkeys are considerable. Neural networks (NNs) are a relatively new option to model growth in animal production systems. One self-organizing sub-model of artificial NNs is the group method of data handling-type NNs (GMDH-type NNs). By means of GMDH-type NNs algorithm a model can be represented as a set of neurons in which different pairs in each layer are connected through a quadratic polynomial and thus produce new neurons in the next layer, therefore can be used to map inputs to outputs. Such a NNs identification p! rocess needs some optimization method to find the best network architecture. For this purposes in this study, aimed to apply the GMDH-type NNs in order to predict caloric efficiency (CE, g of gain/kcal of caloric intake) in tom and hen

turkeys, genetic algorithms are deployed to design the whole architecture of the GMDH-type NNs. Involved effective input parameters in prediction of CE were Age, dietary ME, CP, Met and Lys. Quantitative examination of the goodness of fit for the predictive model was made using R2 and error measurement indices commonly used to evaluate forecasting model. Based on the results of this study and those reported previously (Ahmadi et al. 2007, 2008), the GMDH-type NNs appears a promising method for modeling the relationship between dietary concentrations of nutrients and poultry performance which can be used in choosing and developing special feeding programs to decrease production costs. Also, it can enhance our ability to predict other economic traits, make precise prediction of the nutrition requirements, and achieve optimal performance in poultry production systems.

Key Words: Modeling, group method of data handling-type neural networks, genetic algorithms, caloric efficiency, turkeys

The interplay between social hierarchy and disease dynamics
Davidson, RS, Marion, G and Hutchings, MR

Scottish Agricultural College, Edinburgh, UK, ross.davidson@sac.ac.uk

Heterogeneity in contact patterns arise as a consequence of realistic features of natural populations, and can give rise to significant departures in rates and patterns of disease spread from homogeneous models. Such features can be modelled using a parameter attached to each of a number of subpopulations, with demographic processes being expressed as functions of these parameters. An example is that of social hierarchy, where the parameters can be used to reflect an increase in fitness of dominant individuals. Here we present a model of infection in a hierarchically structured population, and a moment closure technique used to provide approximate analytical solutions for systems of this type. Stochastic, agent based numerical models are compared with analytical results from both first and second order moment closure calculations. From this it is seen that much of the observed prevalence originates from second order effects, and can be attributed to correlations acting in the system arising directly from the presence of hierarchy.

Key Words: Disease dynamics; moment closure

Generalized Non-Homogeneous Semi-Markov and Bernoulli Processes for Salary Lines Construction

Giuseppe Di Biase(*), Fulvio Gismondi(+) and Raimondo Manca(^)

(*)Università "G. D'Annunzio", Chieti-Pescara, Italy, dibiase@unch.it
(+)Università degli Studi Guglielmo Marconi, Roma Italy, f.gismondi@unimarconi.it
(^)Università di Roma "La Sapienza", raimondo.manca@uniroma1.it

Abstract. This paper presents a class of useful models for forecasting in a complete way the salary lines evolution taking into account the salary costs, the number of workers at each rank and the probability transitions among the ranks of a given firm. We speak about a class of models because the salary evolution is function of the career development rules. These rules, usually, change if the observed company changes.
The models are also useful for forecasting the cost of the labour force in the future years and can also be used for following the career development of the workers that will be employed by a given firm in the future. As well known, the salary line forecasting assumes a relevant role in manpower and in the pension fund management..
The model uses the Generalized NHSMP for the construction of the probability promotion and the generalized Bernoulli processes for the construction of salary lines.

Simulation of Multivariate Innovation Diffusion Models

Yiannis Dimotikalis

Department of Finance and Insurance
Technological Educational Institute of Crete
Aghios Nikolaos, Crete, Greece
jdim@staff.teicrete.gr

Innovation diffusion processes typically studied by analyzing univariate time series data. In the case of more than one "suppliers" or products of innovation, the competition can be modeled by multivariate time series data. Discrete multivariate models of innovation diffusion are simulated to exhibit their behavior and ability to represent real time series data.

On classifying coherent/incoherent short texts

Anca Dinu

University of Bucharest, Faculty of Foreign Languages and Literatures, Bucharest, Romania
anca_d_dinu@yahoo.com

In this article we propose a quantitative approach to a relatively new problem: categorizing text as pragmatically correct or pragmatically incorrect (forcing the notion, coherent/incoherent). The typical text categorization criterions comprise categorization by topic, by style (genre classification, authorship identification), by expressed opinion (opinion mining, sentiment classification), etc. Very few approaches consider the problem of categorizing text by degree of coherence, as in (Miller, 2003). One example of application of text categorization by its coherence is creating a spam filter for personal e-mail accounts able to cope with one of the new strategies adopted by spamers. This strategy consists of encoding the real message as picture (impossible to directly analyze and reject by the text oriented classical filters) and accompanying it by a text especially designed to surpass the filter. The role of these e-mail messages is double: to surpass the spam filter so that get to be read by the owner of the account and, second and more important, to untrain the classical spam filter when manually labeled as spam. Thus, after the spam filter sees enough such messages labeled as spams, it eventually cannot make the difference between spam and normal messages. An important question for automatically categorizing texts into coherent and incoherent is: are there features that can be extracted from these texts and be successfully used to categorize them? We propose a quantitative approach that relies on the use of ratios between morphological categories from the texts as discriminant features. We suppose that these ratios are not completely random in coherent text. The goal of our experiment is to automatically classify e-mail messages into two classes: coherent messages, to go to Inbox and incoherent messages (good candidates for Bulk folder). We use supervised machine learning techniques on a small corpus of English e-mail messages and let the algorithms extract important features from all the pos ratios. The results are encouraging (because there are inherent errors, transmitted from the pos tagger and perhaps from the subjective human classification and because using only the pos frequencies disregards many other important feature for text coherence.):

Learning method type	Accuracy
Regression (baseline)	68.18%
linear Support Vector Classifier	78.18%
quadratic Support Vector Machine	81.81%
linear Kernel Fisher discriminant	75.46%
polynomial Kernel Fisher discriminant	85.46%

References (selected)

1. Chih-Chung Chang and Chih-Jen Lin. 2001. LIBSVM: a library for support vector machines. Software available at http://ww.csie.ntu.edu.tw/ cjlin/libsvm.
2. John S. Taylor and Nello Cristianini. 2004. Kernel Methods for Pattern Analysis. Cambridge University Press, New York, NY, USA.
3. Duda, R.O., Hart, P.E., Stork, D.G. 2001. Pattern Classification (2nd ed.). Wiley- Interscience Publication.

A rank based multi-classifier system in text categorization

Liviu P. Dinu

University of Bucharest, Faculty of Mathematics and Computer Science, Academiei 14, RO-0010014, Bucharest, Romania, ldinu@funinf.cs.unibuc.ro

Abstract. In this paper we show that Rank Distance Aggregation can improve ensemble classifier precision in the classical text categorization task by presenting a series of experiments done on a 20 class newsgroup corpus, with a single correct class per document. We aggregate four established document classification methods (TF-IDF, Probabilistic Indexing, Naive Bayes and KNN) in different training scenarios, and compare these results. Keywords: text categorization, rank distance categorization, Reuters database.

THE ASYMPTOTIC BEHAVIOR OF STOCHASTIC RECURRENCE EQUATIONS WITH SYMETRICALLY DISTRIBUTED COEFFICIENTS AND APPLICATIONS FOR ASIAN OPTION PRICING

D. Donchev, M Milev

Sofia University, Sofia, Bulgaria, doncho@fmi.uni-sofia.bg
Plodiv, Bulgaria, marian_milev@hotmail.com

We consider the stochastic equation $Z_t=A_tZ_{t-1}+1$, $t=1,2,…$, A_t being symmetrically distributed i.i.d. random variables. We represent the solution to this equation as an infinite product of random variables that form a Markov chain. We study the relationship between the stationary distribution of this chain and the distribution of A_1. The results permit to be explored the random movement of the asset price when arithmetic Asian options are valuated in the Black-Scholes framework.

Keywords: Stochastic Recurrence Equation, Asian Option Pricing

Comparing Interval-Valued Variables Using S_{LC} Coefficient

Isabel Pinto Doria, Georges Le Calvé and Helena Bacelar-Nicolau

LEAD, Faculty of Psychology and CEAUL, University of Lisbon, Portugal irpdoria@fp.ul.pt
Société Française de Statistique, France, g.lecalve@free.fr
LEAD, Faculty of Psychology and CEAUL, University of Lisbon, Portugal, hbacelar@fp.ul.pt

s_{LC} coefficient is a similarity index between variables proposed by Georges Le Calvé (1977), based on Daniels (1944) and Lerman's (1973) coefficients. In this paper we present the s_{LC} coefficient generalized to compare interval-valued variables, using real data presented by Guru, Kiranagi and Nagabhushan (2004). This symbolic data refers to the minimum and maximum temperatures registered during the twelve months of a certain year in twenty cities. By means of hierarchical cluster analysis and principal component analysis on the similarity matrix obtained with the generalized coefficient s_{LC} we visualize the similarity between the twelve months of the year described by their interval [minimum, maximum] temperatures registered on these cities. This approach permits visualize the structure of proximity of the data, being highly consistent with the knowledge we have about the data.

Keywords: Interval-valued variables, Similarity coefficient, Hierarchical cluster analysis, Principal component analysis, Symbolic data.

References:
1. Daniels, H.E., The relation between measures of correlation in the universe of sample permutations. *Biometrika* **33**, 129-135 (1944).
2. Guru, D.S., Kiranagi, B.B. and Nagabhushan, P., Multivalued type proximity measure and concept of mutual similarity value useful for clustering symbolic patterns. *Pattern Recognition Letters* **25**, 10, 1203-1213 (2004).
3. Le Calvé, G., Un indice de similarité pour des variables de type quelconque. *Revue Statistique et Analyse des Données* **01/02**, 39-47 (1977).
4. Lerman, I.C., *Étude Distributionnelle de Statistiques de Proximité entre Structures Algébriques Finies de Même Type. Application à la Classification Automatique*. Cahiers du Bureau Universitaire de Recherche Opérationnelle, n. 19. Institut de Statistique de l'Université de Paris, Paris (1973).

CUSUM tests based on grouped observations
Karl-Heinz Eger[1] and Evgeni Borisovich Tsoy[2]

[1]Chemnitz University of Technology, Department of Mathematics, Chemnitz, Germany
eger@mathematik.tu-chemnitz.de
[2]Novosibirsk State Technical University, Department of Applied Mathematics, Novosibirsk, Russia, ebcoi@nstu.ru

This paper deals with CUSUM tests based on grouped or classified observations. The computation of average run length is reduced to that of solving of a system of simultaneous linear equations. Moreover a corresponding approximation based on the Wald approximations for characteristics of sequential likelihood ratio tests is presented.
The effect of grouping is investigated with a CUSUM test for the mean of a normal distribution based on F –optimal grouping schemes. The considered example demonstrates that hight efficient CUSUM tests can be obtained for F –optimal grouping schemes already with a small number of groups.
Index Terms: CUSUM, continuous inspection schemes, average run length, grouped observations, classified observations, sequential tests, sequential analysis.

Thermal and respiratory responses of Farafra sheep as affected by some dietary salts under heat stress

M. A. A. El-Barody,* T. M. M. Abdel Khalek**, A.K.I. Abd El-Moty* and A. A. K. Saleh**

*Animal Production Department, Faculty of Agriculture, Minia University.
** Animal production Research Institute, Agriculture Research Center, Ministry of Agriculture.

Thirty five Farafra mature ewes and fifteen rams were used in this investigation. They were divided into five equal groups (15 ewes + 3 rams per each) according to their age and initial body weight as control (G1), 1% sodium bicarbonate (G2), 1% potassium carbonate (G3), 0.5% sodium bicarbonate + 0.5% potassium carbonate (G4) and 1% sodium bicarbonate + 1% potassium carbonate (G5). Ambient temperature, relative humidity and thermal responses (rectal temperature, skin temperature and ear temperature) were measured were recorded simultaneously at 6-8 a.m. and at at 12-2 p.m..(June to August). Moreover, respiratory activates and gas exchange was determined three times (at 15/Jun., 15/Jul. and 15/Aug.) in both rams and ewes throughout experimental period. The results indicated that the dietary salts resulted in a significantly decrease (P<0.01) in rectal temperature (RT), skin temperature (ST) and respiration rate (RR) at 6-8 a.m. and 12-2 p.m. as well as, respiratory quotient (RQ)

in both rams and ewes. On the other hand, Gas volume (GV) and tidal volume (TV) and volume oxygen consumption (VO2) significantly increase (P<0.05). In conclusion,   dietary salts play a great role in alleviation of heat stress on sheep.

Socioeconomic Assessment of Public Pump Irrigated Schemes of the River Nile State in resources use efficiency

Elgilany A. Ahmed[1], Hamid H.M. Faki[2], Hashim A. Elobeid[3]

[1, 2] Ministry of Science and Technology, Agricultural Economics Research and Policy Center, Sudan
[3] University of Khartoum, Faculty of Agriculture, Sudan

The River Nile State (RNS) of Northern Sudan can rightly be considered as a river –born nation that a high population density exists in the settled areas along the River Nile and Atbara River with a total of about 720,000 people (90%), thus, most of agriculture production depends on irrigation in valley areas close to the rivers and they covers approximately 124.000 km^2 (29.5 million feddans) out of which about 3.201.300 fed is suitable for agricultural production. The RNS is one of the areas where large investments in irrigation take place for the production of cash crops. The increased use of natural resources puts a great pressure on the local ecosystem. This fact necessitates an adoption of various techniques of natural resources conservation including the large meaning of natural resources use efficiency. The competition for irrigation water and land increases resource management complexity. The sustainability of irrigated agriculture is questioned and the challenge is to increase simultaneously land and water productivity in the face of the limited availability of land and water in the RNS. The paper undertook Elzeidab irrigated scheme of RNS as study area. The aim of this research is to assess the social and economical performance of Elzeidab scheme tenants and to identify options to improve the local ecosystem system and the available natural resources management. The community based small-scale farms *(Hawasha)* and the option of irrigation is mandatory in Elzeidab from the River Nile (RN) by pumps through the surface irrigation system. To realize these objectives structured survey questionnaires, field observations and literature were used. Various Elzeidab tenant' socio-economic characteristics were gathered to study the socio-economic impact of irrigation water use. A total of 70 randomly selected respondents from Elzeidab scheme were interviewed. Integrated techniques involving economic and hydrologic components are used to assess natural resources use efficiency in RNS.A Descriptive statistics and quantile analysis for crop water applied and crop water requirements for Elzeidab field crops are presented. Stochastic frontier, GAMS, Crop Wat4 and Cobb-Douglas function have been employed to evaluate the social and economical performance of Elzeidab scheme tenants. The results suggest that vast natural resources devoted for agricultural production in the State coupled with low production will need attention on natural resources management, allocation, quantities and introduction of water saving technologies. Water management institutions are not well qualified to handle irrigation water. Lack of tenants' awareness led to inefficient water use. The paper concluded that, to improve the economic and environmental performance of public pump irrigation schemes of the State, numerous challenges are needed to contribute in saving natural resources in the future: institutional support (input supply, output marketing and credit services), training of tenants on improved crop and water management issues, regular supervision and monitoring of scheme activities are crucial.

Neural networks for nonlinear stochastic systems modeling

Ayachi Errachdi*, Ihsen Saad and Mohamed Benrejeb

Research Unit on Automatic (LARA), Department of Electrical Engineering, National School of Engineers of Tunis, Tunisia

Abstract—In this paper, an on-line algorithm using an adaptive learning rate is proposed for the modeling of the multivariable nonlinear stochastic systems. Different cases of Signal -to-Noise Ratio (SNR) are taken to show the effectiveness of this algorithm. The development of an adaptive learning rate is based on the analysis of the convergence of the conventional gradient descent method for the neural networks. A comparative study between the application of a neural networks using a variable and fixed learning rate for multivariable nonlinear stochastic systems is treated. The effectiveness of the proposed algorithm applied to the modeling of behavior of nonlinear dynamic stochastic systems is demonstrated by simulation experiments. The results of simulation showed that the use of neural networks with an adaptive learning rate is more interesting than a fixed learning rate. Two types of non linear stochastic systems are taken.
Key words — multivariable system, nonlinear, stochastic, neural networks, modeling, learning rate.

Call Centers Staffing: A Multicriteria Approach

Jorge Sá Esteves

CIDMA --- Center for Research & Development on Mathematics and Applications, Department of Mathematics, University of Aveiro, Campus de Santiago, 3810-193 Aveiro, Portugal, saesteves@ua.pt

A call center is a service network in which agents provide telephone-based services. Customers who seek these services are delayed in tele-queues. The dimensioning of a large call center is a highly complex problem which implies the extensive use of decompositions of the associated large-scale optimization problems. This is further complicated by the stochastic nature of the teletraffic flows to be handled by the agents. All these requirements lead to a very significant number of interrelated optimization problems and sub-problems which may be approached with different formulations. In this talk we examine the general issue of optimal staffing strategy in the context of call centers for a given stationary input period by period (SIPP). The call center is modeled as a multidimensional Erlang-C system. In practice three main objectives are present in this formulation. Firstly an efficiency criterion requires the total number of waiting calls in the system, to be minimized. On the other hand, a criterion of uniformity in the grade of service measured by the mean waiting time, entails that the absolute difference between the mean waiting times experienced by the calls in the different service groups must be as small as possible (ideally it should be zero). A third criteria is very popular in queueing systems synthesis: the minimization of the proportion of calls which must wait more than a specified limit time. The optimization problem may be seen as a tricriterion convex separable continuous problem. Some mathematical results characterizing the set of the nondominated solutions of the tricriteria formulation, are present. An algorithm for traveling on that set is presented based on the resolution of monocriteria convex problems using Newton-Raphson method, is proposed. Some numerical and graphical results are also presented. Finally, possible advantages of the tricriterion formulation of the problem and its implications on call centers staffing models, are discussed.
Keywords: Stochastic Modeling, Call Centers Staffing, Queueing Systems Dimensioning, Erlang-C System, Multicriteria Optimization.

Some notes on stochastic process foundations and probability distributions in stock market analyses: an interpretive case study of the Nigerian Stock Market

Patrick Oseloka EZEPUE[a] & Mahmoud A. Taib OMAR[b]

[a]Research Leader, Business Intelligence & Quantitative Modelling Research Group, Sheffield Hallam University, Sheffield, S1 1WB, England, UK. Email: p.ezepue@shu.ac.uk
[b] PhD student, Business Intelligence & Quantitative Modelling Research Group, Sheffield Hallam University, Sheffield, S1 1WB, England, UK. Email: mataib2@my.shu.ac.uk

In this paper we review stochastic models and probability distributions typically used in finance and stock market analyses and apply the models and distributions to overall market index and returns data from the Nigerian Stock Market (NSM). The datasets span different financial policy epochs as well as the 2007-09 global financial crises. It is known that stock market data vary from sector to sector whilst being of interest to individual investors, firms and government policy makers. The paper therefore compares results obtained at overall market level to the behaviour of stock market data from the financial services sector which dominates the stock market. Interpretations of the model and distributional results sheds light on the possible impact of different financial policies and reforms as well as the global financial crises on investment and risk management decisions based on such features as market efficiency, bubbles and volatility.
Key Words: Stochastic models, stochastic calculus, probability distributions, stock indexes and returns, investments, risk management

Uncertainty of random variables

Zdeněk Fabián

Institute of Computer Sciences, Academy of Sciences of the Czech republic, Pod vodárenskou věží 2, Praha 8, zdenek@cs.cas.cz

In the first part of the paper we explain the way to an introduction of a scalar score function of continuous random variables with arbitrary support. For some classes of distributions, this function is equal or proportional to the maximum likelihood score for the most important parameter. Overall, it is a new function. Its square, suitably normalized, can be taken as the uncertainty function of the continuous random variables. The expectation of its square can be taken as the generalized Fisher information.
In the second part of the paper, the scalar score function is generalized for discrete random variables. It makes possible to introduce an uncertainty function based on the generalized Fisher information and to compare its mean value with the Shannon entropy.

References
Fabián, Z., ``Estimation of simple characteristics of samples from skewed and heavy-tailed distribution''. In Skiadas, C. (Ed.), Recent Advances in Stochastic Modeling and Data Analysis, World Scientific, Singapore, 43-50, (2007).
Fabián, Z.,`` New measures of central tendency and variability of continuous distributions``. Comm. Statist. Theory Methods, 37, 159-174, (2008).

Coefficient of Variation: Connecting Sampling with some Increasing Distribution Models

Nicolas Farmakis

Dept of Mathematics, Aristotle University of Thessaloniki, Greece

A new use of Coefficient of Variation (Cv) is presented in order to define Distribution Models via Sampling. The interesting case of the simply increasing continuous probability distribution functions (pdf) is studied and the suitably obtained model is presented. A first level checking via sampling and graphs (e.g. histogram) is taking place in order to examine if we have an increasing or no model to be defined. After this using the idea of Cv and some polynomial forms the model is defined. In order to examine if the model is a well approximating one we check the correspondence of the sample data to the expected outputs from the model, by a Chi-square test. So the obtained model becomes a self checked one. The presented distribution models are polynomial and they have a real exponent k>-1. The value k=-1 seems to be a kind of an absolute zero point for the polynomial pdf. The polynomial form of pdf is an approximation to the really existing pdf with more complicated form. This idea was also used in Farmakis (2003) for the symmetric continuous distributions (scd).
Key Words: Sampling, Distribution Models, Coefficient of Variation.

References:
FARMAKIS N. (2001) "*Statistics, Theory in Brief-Exercises*", A & P Christodoulidi, Thessaloniki, 2001 (in Greek).
FARMAKIS N. (2003) "Estimation of Coefficient of Variation: Scaling of Symmetric Continuous Distributions", *Statistics in Transition,* Vol. 6, Nr 1, pp 83-96.
FARMAKIS N. (2007) "*Introduction to Sampling*", Ed. A & P Christodoulidi, Thessaloniki, 2007 (in Greek).

Self-Excited Multifractal Model for Return Fluctuations

Vladimir A. Filimonov and Didier Sornette

Nizhny Novgorod State University, Department of Radiophysics, Nizhny Nogvorod, Russia, vladimir.a.filimonov@gmail.com
ETH Zurich, Chair of Entrepreneurial Risks, Zurich, Switzerland, dsornette@ethz.ch

Scale invariance phenomena of many complex systems could not be characterized by only one intrinsic scale and requires the so-called multifractal description – introducing the continuous spectra of scales. Such multifractal properties are found in a wide range of objects in different domains of science. Started with the models of velocity increments and energy dissipation in developed turbulence, nowadays multifractal models are used to describe such critical systems as triggered seismicity in geophysics, asset return fluctuations in financial markets, traffic flow in computer networks and healthy human heart-beat rhythm in biology.

The first model that exhibit explicit time dependence was the so-called Multifractal Random Walk (MRW) that has been introduced by Bacry, Delour and Muzy. It was the only continuous stochastic stationary causal process with exact multifractal properties and Gaussian infinitesimal increments. However, although this model has no demerits inherent in the preceding models, it looses its sense for high-order incremental moments. The continuous-time Quasi-Multifractal model proposed by Saichev and Sornette and then extended by Saichev and Filimonov was free of this drawback. Being the development of the random walk model, it included several additional significant parameters, which make it possible to avoid the disadvantages of the MRW model.

The generalities of existing models described above are their "exogenous" form, which means that all of them are driven by external noise sources and do not describe the influence of past realization of the process onto to the future values – the mechanism which is essential for such applications as earthquakes and financial markets. In this work we would like to introduce the first "endogenous" multifractal model, which has explicit feedback of past values

onto the future ones. The paper presents the new Self-Excited Multifractal (SEMF) model, discusses its novelty and key features and also describes the properties of the model.

With only three parameters the SEMF model generates the most important stylized facts of empirical financial returns, such as long-range dependence, leverage effect, multifractal scaling of moments and fat tails of the probability distribution. In addition it allows for efficient statistical estimation and provides a natural framework to distinguish between the impact of the exogenous news and the endogenous dynamics.

On Some Modeling Technique for Stochastic Dependences

Jerzy K. Filus and Lidia Z. Filus

Department of Mathematics and Computer Science, Oakton Community College, Des Plaines, Illinois, USA, jfilus@oakton.edu
Department of Mathematics, Northeastern Illinois University, Chicago, Illinois, USA, L-Filus@neiu.edu

We present a powerful, while still simple technique for determining stochastic dependences between various sorts of random variables or random vectors. Application of this technique to a variety of practical problems that one encounters in the 'real world', results in a remarkably wide range of flexible stochastic models.
Keywords: Stochastic modeling. Stochastic dependences, Weak transformations

Brief Description

In this presentation we will restrict the range of applications to the reliability and biomedical areas, where we define a method of modeling stochastic dependences of (any) random **lifetime,** say **Y** from some corresponding random **stresses**

$X_1, \dots ,X_k$, see [1].

The method for such models construction is based on a simple observation that some important stochastic **dependences** of a random quantity Y **[**whose baseline cdf, (i.e., the cdf in an absence of the considered stresses) is F(y; θ) **]** on a set of other **(explanatory)** random quantities $X_1, \dots ,X_k$ can nicely be described by use of a simple, but apparently overlooked in literature notion of "**weak transformations**". These transformations are to be defined on sets of random variables, while valued in 'probability distribution functions' rather than in corresponding to those distribution(s) random variable(s) **Y.** Thus, whenever it fits to a considered practical situation, we can define (and then statistically verify) a particular stochastic dependence by means of a proper weak **transformation that** symbolically may be expressed as:

$(X_1, \dots ,X_k) \rightarrow F(y; \theta)$.

That is in contrast to the ordinary **"strong"** or "direct" **transformation**

$(X_1, \dots ,X_k) \rightarrow Y$.

Among several possible ways the weak transformation can be determined, we have chosen possibly the simplest one by assuming that the **impact** of the (random) quantities (say, "stresses") $X_1, \dots ,X_k$ **on the cdf** of the quantity of interest **Y** exhibits itself as an **impact on** the scalar or vector **parameter** of its probability distribution **F**(y; θ).

Thus, the defined above weak transformation, can be shorten to the relation

$\theta \rightarrow \theta(X_1, \dots ,X_k)$,

where the symbol 'θ' alone, represents the parameter's baseline value i.e., the numerical value in an absence of the stresses $X_1, \dots ,X_k$.

The new (random) value $\theta (X_1, \dots ,X_k)$ of the parameter of Y's cdf may, in the most general case, be considered as just any (parametric) continuous function of its k arguments.

That way yields to explicit determination of the stochastic (weak) dependences in form of the conditional probability distribution of **Y,** given an elementary event

$(X_1, \dots ,X_k) = (x_1, \dots ,x_k)$ realizes.

In symbols that fact can be expressed as:

F(y | $x_1, \dots ,x_k$) = F(y; $\theta(x_1, \dots ,x_k)$).

As it will be shown, such a simple technique yields a huge number of analytically treatable stochastic models for a variety of the real world applications.

For a slightly different, but based on the same idea approach to the considered modeling procedures can be found in [2].

References

1. D. Collett, *Modeling Survival Data in Medical Research*, Chapman @ Hall / CRS A CRC Press Company, Second Edition, New York, Washington , D. C. (2003)

2. J. K. Filus, L. Z. Filus, B. C. Arnold, <u>Families of Multivariate Distributions Involving "Triangular" Transformations</u>, *Communications in Statistics -Theory and Methods*, Volume <u>39</u>, Issue 1, 107-116 (2010)

Confidence bands for comparing percentile residual life functions

Alba M. Franco-Pereira, Rosa E. Lillo and Juan Romo

Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain, alba.franco@uc3m.es
Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain, rosaelvira.lillo@uc3m.es
Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain, juan.romo@uc3m.es

The percentile residual life orders were introduced in Joe and Proschan (1984) and extensively studied in Franco-Pereira, Lillo, Romo, and Shaked (2009). In this paper, some interpretations and properties of these stochastic orders were given and some applications in reliability theory and finance were described.
Given the advantages of the percentile residual life orders, especially in practical situations, it is convenient to develop an statistical tool to test whether two independent random samples have underlying random variables which are close with respect to a γ-percentile residual life order.
In this work, we present a nonparametric method for constructing confidence bands for the difference of two percentile residual life functions. This functional data analysis technique incorporates bootstrap resampling and the concept of statistical depth. The confidence bands provide us with evidence of whether two random variables are close with respect to some percentile residual life order. The practical performances of the bands are evaluated through simulation. Some applications with real data are also given.
Keywords: Confidence bands, functional data analysis, bootstrap, percentile residual life, statistical depth.

NEW INSIGHTS FOR RECURSIVE RESIDUALS

James M Freeman

Manchester Business School, University of Manchester, Booth Street West
Manchester, M15 6PB, UK, jim.freeman@mbs.ac.uk

The role of recursive residuals in the validation of linear models is well-recognised. Furthermore, their contribution to quality control and change point analysis has become increasingly valued. A new procedure for deriving recursive residuals – assuming data observations are time-ordered - is now presented and its theoretical efficacy established. Relevant results are illustrated across a number of real-life data applications.
Keywords: Influence, Leverage, Outliers

Optimal Design of Smoothing Spline Curves with Equality and Inequality Constraints

Hiroyuki Fujioka[1] and Hiroyuki Kano[2]

[1] Department of System Management, Fukuoka Institute of Technology, 3-30-1,Wajiro-higashi,Higashi-ku, Fukuoka 811-0295 Japan, fujioka@fit.ac.jp
[2] School of Science and Engineering, Tokyo Denki University Hatoyama, Hiki-gun, Saitama 350-0394, Japan, kano@mail.dendai.ac.jp

In this paper, we consider the problem for designing of optimal smoothing spline curves with equality and/or inequality constraints. The splines are constituted employing normalized uniform B-splines as the basis functions. Then various types of constraints are formulated as linear function of the so-called control points, and the problem is reduced to convex quadratic programming problem. The performance is examined by some numerical examples.
Keywords: B-splines, optimal smoothing splines, equality/inequality constraint, quadratic programming.

Patient pathway prognostication using the extended mixed distribution survival tree based analysis

Lalit Garg[1], Sally McClean[1], Maria Barton[1], Brian Meenan[2] and Ken Fullerton[3]

[1] School of Computing and Information Engineering, University of Ulster, Coleraine, UK
garg-l@ulster.uk, si.mcclean@ulster.uk, m.barton@ulster.uk
[2] School of Engineering, University of Ulster, Jordanstown, UK, bj.meenan@ulster.ac.uk
[3] College of Medicine and Health Sciences, Queen's University, Belfast, UK,
ken.fullerton@belfasttrust.hscni.net

Mixture distribution survival trees are constructed by approximating different nodes in the tree by distinct types of mixture distributions to achieve more improvement in the likelihood function and thus the improved within node homogeneity. In our previous work (Garg et al. 2009), we proposed a mixed distribution survival tree based method (where tree nodes were approximated using Gaussian mixture distributions and phase type distributions) for determining clinically meaningful patient groups from a given dataset of patients' length of stay where partitioning was based on covariates representing patient characteristics such as gender, age at the time of admission, and primary diagnosis code. This paper extends this approach to patient pathway prognostication i.e. for determining importance and effects of various input covariates such as gender, age at the time of admission and primary diagnosis code on patients' hospital length of stay and to examine the relationship between length of stay in hospital and discharge destination (or treatment outcome). An application of this approach is illustrated using 5 year retrospective data of patients admitted to Belfast City Hospital with a diagnosis of stroke (hemorrhagic stroke, cerebral infarction, transient ischaemic attack TIA, and stroke unspecified).
Keywords: Stochastic modeling, Survival tree, Length of stay modelling, Prognostication, Clustering, Gaussian mixture distributions, Phase type distributions

References
-Garg, L., McClean, S. I., Meenan, B. J., El-Darzi, E. and Millard, P. H., Clustering patient length of stay using mixtures of Gaussian models and phase type distributions. *The 22nd IEEE Symposium on Computer-Based Medical Systems (CBMS 2009),* Albuquerque, New Mexico, USA, August 3-4, 2009, pp. 1-7, doi:10.1109/CBMS.2009.5255245 (2009).

On Katz type recursions with applications to insurance

Maude GATHY and Claude LEFEVRE

Université Libre de Bruxelles, Bruxelles, Belgium, mgathy@ulb.ac.be

The classical Katz recursion is a first-order recursion of simple form depending on two parameters (Katz [3]). It is known to characterize the binomial, Poisson and negative binomial distributions. In insurance, its main usefulness is to yield a simple recursive algorithm for the aggregate claims distribution (Panjer [4]).

The present work is first concerned with the Markov-Polya distribution and its links with the Katz family of distributions. The Markov-Polya distribution is presented as a claim frequency model incorporating some (anti)contagion effects.

This distribution is shown to satisfy a Katz-like recurrence and this enables us to derive a simple recursion that generalizes the Panjer algorithm. The Katz family is also obtained as a limit of the Markov-Polya distribution. An observed frequency of car accidents is fitted and shows the suitability of the Markov-Polya distribution.

The paper then focused on the Lagrangian Katz family of distributions which also satisfies an extended recursion depending this time on three parameters. The distributions covered by the recursion are shown to be the generalized Poisson, generalized negative binomial and binomial distributions. The index of dispersion, an extended Panjer algorithm and the asymptotic tail behaviour of the family are then presented. This recursion is also derived through a first-crossing problem which models two applications in risk theory. Finally, the relevance of the family is illustrated with several data sets on the frequency of car accidents.

References
1. Gathy, M. and Lefèvre, Cl., "On Markov-Polya distribution and the Katz family of distributions", Communications in Statistics: Theory and Methods, to appear (2009).
2. Gathy, M. and Lefèvre, Cl., "On the Lagrangian Katz family of distributions as a claim frequency model", submitted (2009).
3. Katz, L. , "Unified treatment of a broad class of discrete probability distributions", in Classical and Contagious Discrete Distributions, Pergamon Press, Oxford, 175-182 (1965).
4. Panjer, H. H., "Recursive evaluation of a family of compound distributions", ASTIN Bulletin, 12, 22-26 (1981).

Key Words: Katz recursion, Katz family of distribution, Markov-Polya laws, Lagrangian Katz family of distributions, aggregate claims distribution, extended Panjer algorithm, limit distribution, asymptotic tail behaviour

---

Differentiation Tests for Three Dimensional Shape Analysis

Stefan Giebel, Jens-Peter Schenk and Jang Schiltz

Luxembourg School of Finance, Luxembourg, stefan.giebel@uni.lu
Paediatric Radiology, University of Heidelberg, Germany, jens-peter.schenk@med.uni-heidelberg.de
Luxembourg School of Finance, Luxembourg, jang.schiltz@uni.lu

There are different kinds of tumours in childhood: nephroblastoma, clear cell sarcoma, neuroblastoma etc. The chosen therapy depends upon the diagnosis of the radiologist which is done with the help of MRI (Magnetic resonance images). Our research is the first mathematical approach on MRI of renal tumours (n=80). We are using transversal, frontal and sagittal images and compare their potential for differentiation of the different kind of tumours by use of Statistical Shape Analysis.

Statistical shape analysis is a methodology for analyzing shapes in the presence of randomness. It allows to study two- or more dimensional objects, summarized according to key points called landmarks, with a possible correction of size and position of the object. So

objects with different size and/or position can be compared with each other and classified. To get the shape of an object without information about position and size, centralisation and standardisation procedures are used in some metric space. This approach provides an objective methodology for classification whereas even today in many applications the decision for classifying according to the appearance seems at most intuitive.

We determine the key points or three dimensional landmarks of the renal tumours by using the edges of the platonic body (C60). We present a new test for the mean shape based on the variance within tumour groups with the same diagnosis. Unlike the classical test from Ziezold (1994) we do not need any more a mean shape in each case for both groups is necessary for differentiation.

Moreover, we apply Logistic regression and Configuration Frequency Analysis for classification on the sample. While Logistic regression handles the data in a continuous type, Configuration Frequency Analysis uses discrete variables.

Eventually we discuss the consequences of our results for the application in oncology.

Keywords: Shape, Mean Shape, Variance, Reflection, Renal tumours

References
1. Dryden, I.L; Mardia, K.V., Statistical Shape Analysis. Wiley, Chichester, 1998.
2. Giebel, S; Schenk, J.P.; Schiltz, J., Analysis on two dimensional objects in medicine: Differentiation of tumours in early childhood. In: Proceedings of the the 20th international congress of Jangjeon Mathematical Society, 2010.
3. Ziezold, H., Mean Figures and Mean Shapes Applied to Biological Figure and Shape Distributions in the Plane, *Biometrical Journal* **36**, 491-510, (1994).

Forecasting Broadband Penetration in Greece using a Stochastic Logistic Growth Model with Dynamic Carrying Capacity

Apostolos Giovanis and Christos Frangos

Technological Educational Institute of Athens, Faculty of Economy and Business Department of Business Administration, Ag. Spiridonos Street, Aigaleo, P.C.122 10, Athens, Greece
giovanis@teiath.gr, cfragos@teiath.gr

With huge investments required to launch and develop broadband services, being able to accurately forecast subscriber take-up and revenues is very important task for the businesses being involved. An accurate forecast of asset required and costs is also critical to building an effective business case, to profitably pricing the services.

This paper follows a stochastic approach to model and forecast the diffusion process of broadband services in Greece. Originating from the well known logistic growth function the authors extend the model to consider both the stochastic nature of the innovation propagation mechanism applied to the growth coefficient of the model and the impact of service's functionality changes, which alters the carrying capacity or creating a new carrying capacity in the process. For the estimation of the proposed stochastic diffusion model, the Augmented Kalman Filter with continuous state and discrete observations (AKF(C-D)), is used. This method is directly applicable to any differential diffusion model without imposing constraints on the model structure or the nature of the unknown parameters. Furthermore, it provides a systematic way to incorporate prior knowledge about the likely values of unknown parameters, and updates the estimates when new data become available.

The proposed model and its ancestors are compared using their goodness of fit and forecasting accuracy indices to quarterly historical data over a period of 6 years. The comparison revealed that the proposed model is a very appropriate candidate for forecasting broadband diffusion in Greece.

Keywords: Technological innovation and diffusion, Stochastic Logistic growth, Process functionality, Dynamic carrying capacity, Forecasting, Broadband services

Universal fluctuations of the Dow Jones

Rui Gonçalves, Helena Ferreira and Alberto Pinto

LIAAD-INESC Porto LA and Faculty of Engineering, University of Porto, Portugal,
rjasg@fe.up.pt
LIAAD-INESC Porto LA, University of Porto, Portugal,
helenaisafer@gmail.com
LIAAD-INESC Porto LA and Department of Mathematics, Faculty of Sciences,
University of Porto, Portugal
aapinto1@gmail.com

We compute the analytic expression of the probability distributions FDJ,+ and FDJ,− of the normalized positive and negative Dow Jones index daily returns r(t). Furthermore, we define the α re-scaled Dow Jones daily index positive returns r(t)^α and negative returns (−r(t))^α that we call, after normalization, the α positive fluctuations and α negative fluctuations. We use the Kolmogorov-Smirnov statistical test, as a method, to find the values of α that optimize the data collapse of the histogram of the α fluctuations with the Bramwell-Holdsworth-Pinton (BHP) probability density function. The optimal parameters that we found are $\alpha^+ = 0.46$ and $\alpha^- = 0.45$. Since the BHP probability density function appears in several other dissimilar phenomena, our result reveals a universal feature of the stock exchange markets.

Keywords: Data Analysis, Stock Exchange Markets, Universality

Likelihood analysis of errors-in-variables models

Annamaria Guolo

Department of Economics, Society and Institutions, University of Verona, Via dell'Artigliere, 8, Verona, I-37129 Italy, annamaria.guolo@univr.it

The problem of errors affecting the measure of variables is very common in many areas of research, such as epidemiology, biology, health economics and econometrics. It is well known that ignoring the presence of measurement errors in inferential procedures may lead to noneglectable consequences, as for example a substantial bias of the parameter estimators.
In this work we focus on a likelihood-based approach to correct for measurement error affecting covariates in regression models. The likelihood-based approach, while being advantageous with respect to alternative solutions in terms of properties of the parameter estimators, is still less popular in applications, mainly as a consequence of substantial computational efforts. To alleviate this problem, we explore a computational procedure relying on a Monte Carlo methodology. The finite sample performance of the likelihood approach is investigated through simulation studies, and compared to regression calibration, that is currently the most popular measurement error correction technique.
Key Words: maximum likelihood estimator; measurement error; Monte Carlo method; nondifferential error; regression calibration

BIRTH, DEATH - PROCESSES AND THEIR APPLICATIONS TO DEMOGRAPHIC PROBLEMS

KUNASEKARAN GOPALAKRISHNAN

PERIYAR UNIVERSITY, SALEM, TAMILNADU, INDIA, gkunasekaran@yahoo.co.in

Stochastic point process models are widely used in every domain of application where a sequence of times - each time corresponding to the occurrence of some events - constitute the observable data. Vital statistics are data on the occurrence of fundamental events of

human lives such as birth, death, marriage, and the like. Brillinger (1986) discusses the demographic problems through the general properties of point processes. The observation of Brillinger (1986) is that a Poisson property of the population birth process enhances a two dimensional "Poissonness" of the population death process. He argues that, assuming Poisson birth times and independent life times, the number of deaths and the corresponding mid year population in an open population have a bivariate Poisson distribution.

However both fertility and mortality rates differ, in general, between the two sexes. It is unsatisfactory in a stochastic model to have to ignore such a chance departure. Thus it is more appropriate to represent the birth times by a bivariate point process instead of a univariate point process such as Poisson process. In this paper, assuming Poisson birth times and independent lifetimes, it is shown that the death process is a bivariate planar Poisson process (see,for example, Daley and Vere-Jones (1988) for planar point processes). Further, the concept of two dimensional vital rates has been introduced and their joint probability distributions are obtained.

References

[1] Brillinger,D.R.(1986).The natural variability of vital rates and associated statistics (with discussion). Biometrics 42, 693-724.
[2] Daley,D.J. and Vere-Jones,D.(1988).An introduction to the theory of point processes. Springer, London.

Key Words: Point processes, Poisson processes, Vital rates

Stochastic groundwater modeling for a fractured aquifer in Augusta area (Italy)

Enrico Guastaldi, Andrea Carloni, Claudio Gallo, and Vincenzo Ferrara

Centre for GeoTechnologies of University of Siena, San Giovanni Valdarno (AR), Italy
guastaldi@unisi.it, carloni@unisi.it
CRS4 Cagliari, Technology Park Pixinamanna, Pula (CA), Italy, fender@crs4.it
Earth Sciences Department, University of Catania, Catania, Italy, ferrara.vincenzo@eutelia.com

Simulation of groundwater flow in fractured aquifers modeling involves problems related to heterogeneity of the medium. Therefore, natural system needs mathematical approximations. In such aquifer systems, the study of groundwater pollution scenarios and risk analysis are complex because the difficulty in determining both times and the maximum distances covered by pollutants.

The objective is to assess the risk of contamination of deep groundwater in carbonate aquifer of Augusta coastal area (close to Sicilian Syracuse town, in Italy), to evaluate possible aquifer pollution scenarios. In order to reconstruct the complex geological framework of the study area, we utilize a geostatistical approach out of the usual numerical codes. We implemented a numerical model by a stochastic approach to simulate both the groundwater flow (MODFLOW-2000 code) and the pollutants transport (MT3DMS code). A critical step was the regionalization of hydrodynamic parameters: hydraulic conductivity requested probabilistic analyses to minimize uncertainty of spatial distribution of available data.

We assigned the hydraulic conductivity values at the model by implementing an algorithm: we generated a stochastic distribution Monte Carlo type, based on Gaussian probability density function appropriately formulated. Afterwards, we assigned to the model a discrete set of hydrodynamic conductivity values related to the main geological discontinuities.

The model allowed to reproduce complex hydrogeological systems, and then to perform the contamination risk analysis. Furthermore, the algorithm that we programmed represents a computational element that makes the model more flexible. In fact, this code generates different configurations of hydrodinamic parameters for stochastic simulation, the more precautionary way to formulate risk analysis in terms of probability.

Keywords: Stochastic approach, Numerical modeling, Monte Carlo simulation, Risk analysis, Geostatistics, Carbonate aquifer, Groundwater, Sicily

A Note on the Bootstrap Variant of AIC for Periodic State-Space Model Selection

Hafida Guerbyenne and Fayçal Hamdi

Faculty of Mathematics, USTHB, BP 32. El Alia, Bab Ezzouar, Algiers, Algeria.
hamdi_fay@yahoo.fr, hguerbyenne@yahoo.fr

This note extends the bootstrap variant of AIC for time-invariant state-space model selection due to Cavanaugh and Shumway (1997) to the case of periodic time-varying state-space models. We first derive an algorithm which extends the bootstrap procedure proposed by Stoffer and Wall (1991) in the context of Gaussian time-invariant state-space model to the periodic state-space framework. As a result, the derived algorithm is exploited in estimation of the expected Kullback-Leibler discrepancy between the fitted approximating periodic model and the generating periodic model to evaluate the bootstrap variant of AIC. We compare the performance of our AIC variant to that other selection criteria in simulation studies based on small sample sizes.

Keywords: Periodic state-space models, periodic ARMA models, information theory, Kullback-Leibler, bootstrap, finite sample distributions, periodic Chandrasekhar filter.

Ensemble Methods of Appropriate Capacity for Multi-Class Support Vector Machines

Yann Guermeur

LORIA-CNRS, Campus Scientifique, BP 239, 54506 Vandoeuvre-l`es-Nancy Cedex, France
Yann.Guermeur@loria.fr

Roughly speaking, there is one single model of pattern recognition support vector machine (SVM), with variants of lower popularity. On the contrary, among the different multi-class SVMs (M-SVMs) published, none is clearly favoured. Although several comparative studies between M-SVMs and decomposition methods have been reported, no attention had been paid so far to the combination of those models. We investigate the combination of M-SVMs with low capacity linear ensemble methods that estimate the class posterior probabilities.

Keywords: Ensemble methods, M-SVMs, Capacity control.

Analysis of the salary trajectories in Luxembourg: a finite mixture model approach

Jean-Daniel Guigou, Bruno Lovat and Jang Schiltz

Luxembourg School of Finance, Luxembourg, stefan.giebel@uni.lu
University Nancy II, Nancy, France, bruno.Lovat@univ-nancy2.fr
Luxembourg School of Finance, Luxembourg, jang.schiltz@uni.lu

We analyze the salaries of about 700.000 employees who worked in Luxembourg between 1940 and 2006 with the aim of detecting groups of typical salary trajectories with respect to some covariants like sex, workstatus, residentship and nationality. We use the proc traj SAS procedure from Bobby L. Jones to classify the workers and descriptive statistical methods like the CHAID procedure to get a caracterization with respect to the covariants of the different groups.

In a second part of the talk we use these results to analyse the long term sustainability of the Luxembourg pension system. Finally, we compare the current pure repartition model to a possible mix of repartition and capitalization.

More precisely, we propose a probabilistic economic model to decide for each of the groups found above which part of the pension should be financed by repartition and which part by capitalization. It is based on a portfolio type risk management approach to compare the demographic risk inherent to the repartion strategy to the financial risk inherent to the capitalization strategy.

Estimating Unconstrained Demand Rate Functions using Customer Choice Sets

Alwin Haensel, Ger Koole

Department of Mathematics, VU University Amsterdam, De Boelelaan 1081A, Amsterdam, Netherlands, ahaensel@few.vu.nl

A good demand forecast should be at the heart of every Revenue Management model. Yet most demand models focus on product demand and do not incorporate customer choice behavior under offered alternatives. We are using the ideas of customer choice sets to model the customer's buying behavior. A customer choice set is a set of product classes representing the buying preferences and choice decisions of a certain customer group. We are presenting a demand estimation method for these choice sets. The procedure is based on the maximum likelihood method and to overcome the problem of incomplete data or information we additionally apply the expectation maximization (EM) method. Using this demand information per choice sets, the revenue manager obtains a clear view of the underlying demand. In doing so, the sales consequences from different booking control actions can be compared and the over all revenue be maximized.
Key Words: Revenue management, demand forecasting, data unconstraining, customer choice, choice sets, maximum likelihood estimation, EM algorithm.

Recursion for marginals and normalizing constant for Gibbs processes

Hardouin Cécile, X. Guyon

SAMM, Université Paris 1, and MODAL'X, Université Paris Ouest, France

Usually, the computation of the marginal distributions and/or the normalizing constant C of a discrete probability distribution π involves high dimensional summation, such that the direct evaluation of these sums becomes quickly infeasible in practice. For example, for an Ising model on a 10×10 grid, it involves summation over 2^{100} terms. This problem has a deep impact for many applications, for instance, maximum likelihood parameter estimation. Some significant efforts have been put to solve the problem and the litterature displays various alternatives for distributions involving such unreachable constants. However, it is sometimes to compute directly the true marginals and normalizing constant. Here, we first summarize classical results on the calculus of C and of the marginal distributions of π, a general Gibbs distribution. Then, we propose a new algorithm based on π's conditional probabilities to compute the marginals. Particularly, we evaluate the numerical performance for a spatial Ising model on a lattice m×T.
Key words: Gibbs process, Normalizing constant, Ising model

Trends in moving-seasonal time series of temperature and precipitation time series in the Czech Republic

Karel Helman

University of Economics, Prague, Czech Republic, helmank@vse.cz

This paper focuses on time series of the selected climatology elements measured in the Czech Republic territory. It represents a continuation of some previous author's works. The paper concentrates on the past developments. It aims to give a well-founded view on

47

the seasonality development in selected time series. There are two main objectives of the paper. First: the construction of time series composed of moving-seasonal factors calculated from the input time series. Second: the calculation of trends included in the moving-seasonal time series.

Key Words: seasonality, trends, climatology, time series, Czech Republic

On Heat Polynomials and First Hitting Times of Diffusions

Gerardo Hernandez-del-Valle

Columbia University

In this talk we will derive sharp bounds to the first hitting time of general diffusions by first relating the problem to a functional of 3-dimensional Bessel bridge. Furthermore, we will relate the exact densities to series of the so-called heat polynomials.

Random effects models for discrete time-course data

John Hinde

Science Foundation Ireland funded BIO-SI Project, School of Mathematics, Statistics & Applied Mathematics, National University of Ireland, Galway, Ireland.

The random effect and mixed modelling methodology for normally distributed data has been developed extensively in the last twenty plus years and is well understood. While models have been extended to other types of response variable, mainly through generalized linear mixed models for count and proportion data, the behaviour of some aspects of these models is less well known. Software exists for model fitting, but how robust are analyses to various forms of model misspecification? This talk will address aspects of this in the context of analysing some forms of discrete longitudinal data.

For illustration, we will consider data from a toxicological experiment for biological pest control, where the cumulative mortality in groups of insects measured at various time points. Samples of different strains, or isolates, of the compound under study are used, typically with some replication. Cumulative multinomial models provide an obvious approach to the analysis of these data, however, the basic model needs to be extended to account for overdispersion. Here we consider various random effect models, including the Dirichlet-multinomial and models with random effects in the linear predictor. For some simple time trend models, fitting using a generalized estimating equation approach leads to a surprisingly simple modification of the multinomial fit and casts light on the behaviour in more complex models. We will place this result in context of similar behaviour for standard mixed models with special simple variance structures and consider extensions to other settings. These models allow us to address issues of robustness of inference for unit level covariates in longitudinal models to variation over time, and more generally for inference on between cluster covariates.

We will also present some results on the effect of specific assumptions about the random effect distribution, considering the use of flexible random effects, such as the nonparametric approach with its resulting discrete estimate of the random effect distribution, and also the so-called semi-nonparametric and smooth nonparametric approaches that do not make strong assumptions about the shape of the random effect distribution, but which allow some control over the smoothness of the resulting estimate.

A semi-Markov regime switching extension of the Vasicek model

Julien S. Hunt and Pierre Devolder

Université Catholique de Louvain, Institute of Statistics, Louvain-La-Neuve, Belgium,
julien.hunt@uclouvain.be
Université Catholique de Louvain, Institut of Actuarial Sciences, Louvain-La-Neuve, Belgium,
pierre.devolder@uclouvain.be

In this paper, we briefly recall some basic definitions coming from the theory of interest rates. We then define a sound mathematical framework for our model. The idea is to use the power of stochastic calculus. For this, we provide a marked point process approach to semi-Markov processes. With this, we can easily generalize the Vasicek model of the short rate of interest. Our model is a semi-Markov regime switching extension of the Vasicek model. This means we allow the parameters of the model to vary in time according to an underlying semi-Markov process. Some properties of this model are given. We then discuss the notion of risk neutral measures in this context. We move on to study the pricing of interest rate derivatives in our framework. Specifically, we show that the price of a zero-coupon bond has to satisfy a system of integro-differential equations that are influenced both by the market price of risk and by the market price of regime switches.

Functional approach to cluster municipalities with respect to air quality assessment

Rosaria Ignaccolo, Stefania Ghigo and Stefano Bande

Dipartimento di Statistica e Matematica Applicata, Università degli Studi di Torino, Torino, Italy, ignaccolo@econ.unito.it, ghigo@econ.unito.it
ARPA Piemonte, Torino, Italy, s.bande@arpa.piemonte.it

The European Directive 1996/62/CE imposes land classification, to distinguish zones which need further actions from those which only need maintenance. We propose to classify a region through a functional approach on air pollutant time series. We aggregate time series at municipality scale of several pollutants in an air quality index and, alternatively, we summarize them by Functional Principal Component Analysis. Then we partition municipalities and compare these two different strategies.
Keywords: Functional Data Analysis, B-splines, Cluster Analysis, PAM, Atmospheric Pollution.

Multiple Model Adaptive Control with Mixing: Discrete-time Case

Petros Ioannou[1], Edoardo Mosca[2], and Simone Baldi[2]

[1] Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA, ioannou@usc.edu
[2] Dipartimento Sistemi e Informatica, Universit`a di Firenze, 50139 Firenze, Italy, mosca@dsi.unifi.it, sbaldi@dsi.unifi.it

The main result of this paper is the extension of the AMC approach exposed in [1], to the discrete-time setting. Besides, extension to the tracking problem is considered. The stability and robustness properties of the adaptive mixing control scheme are anaylized. It is shown that in the ideal case, when no disturbances or unmodelled dynamics are present, the tracking error converges to zero; otherwise the mean-square tracking error is of the order of the modeling error provided the unmodeled dynamics satisfy a norm-bound condition.
Keywords: Robust adaptive control, Multiple model adaptive control.

Stochastic Cash Flows and Continuous Time Homogeneous and Non Homogeneous Backward Semi-Markov Reward Processes

Jacques Janssen, Raimondo Manca* and Ernesto Volpe di Prignano

*Università di Roma "La Sapienza" , Facoltà di Economia, via del Castro Laurenziano, 9 Roma 00161, Italy, raimondo.manca@uniroma1.it

In this paper, a full treatment of continuous time homogeneous and non-homogeneous backward semi-Markov reward processes will be presented, as far as the authors know, for the first time in the continuous time continuous state non-homogeneous case.

In the continuous time semi-Markov process environment, the distribution function that rules the transitions between the states of the studied system can be of any type and not only exponential. This fact is an important generalization as regards the Markov environment.

The introduction of backward time makes it possible to consider the instant in which the system entered a state, even if it entered before the time under consideration.

Rewards permit the introduction of a financial environment into the model.

Considering all these properties any stochastic cash flow can, in the authors' opinion, be naturally modeled by means of semi-Markov reward processes. Furthermore, the backward case offers the possibility of considering the duration of an event that began before the time in which the system is observed, and this fact can be very useful in the evaluation of some insurance contracts.

 Key Words: semi-Markov, Cash flows, Backward time

On geometry and scale of a stochastic chemostat

Marc Joannides and Irène Larramendy-Valverde

I3M, University of Montpellier II, France

A chemostat is a fixed volume bioreactor in which microorganisms are grown in a continuously renewed liquid medium. We propose a stochastic model for the evolution of the concentrations in the single species and single substrate case. It is obtained as a diffusion approximation of a pure jump Markov process, whose increments are comparable in mean with the deterministic model. A specific time scale, related to the noise intensity, is considered for each source of variation. The geometric structure of the problem, usable by identification procedures, is preserved both in the drift and diffusion term. We study the properties of this model by numerical experiments.

Estimation of innovation diffusion models with autoregressive error term

Juho Kanniainen[1], Arto Luoma[2], Henri Suur-Inkeroinen[1], and Robert Piché[3]

[1] Tampere University of Technology, Department of Industrial Management, P.O. 541, FI-33101 Tampere
[2] University of Tampere, Department of Mathematics, Statistics and Philosophy, FI-33014 Tampere
[3] Tampere University of Technology, Department of Mathematics, P.O. 533, FI-33101 Tampere

The existing literature on innovation diffusions has so far dealt with the problem of selecting a model mostly by assuming that the error term is independent and identically distributed to

produce favorable fit statistics or, at best, good forecasting capability. In this paper we consider the estimation of innovation diffusions by specifying the error process of sales as an AR(n) process. We propose an estimation procedure and analyze our error model empirically with a large set of data. We also show that characterizing the error term as an autoregressive process can greatly impact the performance of forecasting.

On Generalized Divergence Measures and Their Applications in Modeling and Testing Procedures

Alex Karagrigoriou and Kyriacos Mattheou

Department of Mathematics and Statistics, University of Cyprus, Nicosia
alex@ucy.ac.cy, mattheouk@cytanet.com.cy

Divergence measures are used as indices of similarity or dissimilarity between populations. They are also used either to measure mutual information concerning two variables, to perform tests or to construct selection criteria. In this paper we propose testing procedures and selection criteria based on the BHHJ measure of divergence.
Keywords: Measures of divergence, goodness of fit tests, statistical modeling, autoregressive process.

1 Introduction
Measures of discrepancy or divergence as a way to evaluate the distance (divergence) between any two probability distributions or functions have a long history. The most well known family of measures of divergence is the Csiszar's family known also as Csiszar's φ-divergence (Csiszar, 1963). On the other hand one of the most recently proposed measures of divergence is the Basu, Harris, Hjort, and Jones (BHHJ) power divergence which has been recently generalized by Mattheou et. al (2009).
In this paper we focus on the discrete version of the new generalized BHHJ family of measures of divergence and investigate its implementation in testing statistical hypotheses for multinomial distributions. Furthermore, we investigate its implementation in statistical modeling and prediction for the determination of the order in vector autoregressive processes. Simulations are performed to check the appropriateness of the proposed statistics.

References
1. Mantalos, P., Mattheou, K., Karagrigoriou, A. (2010). Forecasting ARMA Models: A comparative study of information criteria focusing on MDIC, *J. Statist. Comput. and Simul*., DOI: 10.1080/00949650802464137 (to appear)
2. Mattheou, K., Lee, S. and Karagrigoriou, A. (2009). A model selection criterion based on the BHHJ measure of divergence, *J. Statist. Plan. Infer*., Vol. **139**, pp. 128-135.
3. Vonta, F. and Karagrigoriou, A. (2010). Generalized measures of divergence in survival analysis and reliability, *J. of Applied Prob*. (to appear).

Estimation of Parameters of SDE. Black- Scholes Model Share Price of Gold

Khaled Khaldi and Samia Meddahi

Department of Mathematics, Faculty of Sciences, Boumerdes University, Boumerdes, Algeria
kkhaldi@umbb.dz
Department of Mathematics, Faculty of Sciences, Boumerdes University, Boumerdes, Algeria
samia.meddahi@gmail.com

The study's most recognized in the world of mathematical finance is certainly the model of option pricing Black-Scholes (1973), which experienced immediate sucked from researchers and financial professionals, as was the need great tool for managing risk. The Black-Scholes model used to evaluate the price of an option in the event of failure arbitration, based on the assumption that the evolution of a title follows a geometric Brownian motion and the volatility

is constant. The Black-Scholes is undoubtedly the first theoretical model of evaluation have been used so intensively by professionals for evaluation, speculation or simply hedging. However, the first problem, which it faces, and for models in general is that of parameter estimation.

We present the Black-Scholes model, it discusses in detail the various techniques for estimating parameters based on the likelihood function, the method "discreet" considers the density function of the transition process log normal distribution. The second proposes the estimation of model parameters through observation of the first passage time of the process through a terminal constant whose density is known. One treats an application of the share price of gold.

Keywords: Estimate parameters, Discrete method, Time of the first passage, Stochastic differential equation

References

1. Black F., Scholes M., The Pricing of Options Liabilties. *The Journal of Political Economy*, Vol.81, N°3, 637-654 (1973).
2. Gross P., Parameter Estimation for Black-Scholes Equation. *Springer* (2006).
3. Janssen J., Saib M., Khamiss T., Techniques d'Estimation pour le Modèle de Black et Scholes. *AFIR Colloquium* Nurnberg, Germany (1996).
4. Lamberton D., Lapeyre B., Introduction to Stochastic Calculus Applied to Finance. *Springer*. (2007).
5. Peirano E., Talay D., Domain Decomposition by stochastic Methods. Fourtheen International Conference on Decomposition Methods. *Editors Herrera I., Keyes D. E., Widlund O. B., Yates R.*, 131-147 (2003).

Estimating Functions for nonstationary ARCH process

Sahm Kim

Department of Statistics, Chung- Ang University, Seoul, Korea, 156-756

Jensen and Rahbek(2004,2006) investigated asymptotic properties of the quasi-maximum likelihood(QMLE) estimator in the nonstationary ARCH and GARCH models. In this article, we first introduce the estimating function based on the Godambe's (1985) criterion and establish the consistency and asymptotic normality of the estimator which is the solution of the proposed estimating equation for the nonstationary ARCH process. Performance evaluation between QMLE and the proposed estimator is shown by simulation study.

Keywords: ARCH, consistency, asymptotic normality, estimating function, quasi-maximum likelihood estimation, nonstationarity

The Sensitivity of Sequential Triple Sampling to Departures from Normality

Alan Kimber and Ali Yousef

Southampton Statistical Sciences Research Institute, University of Southampton, UK
A.C.Kimber@soton.ac.uk
School of Mathematics, University of Southampton, UK, alisyousef@yahoo.com

In 1981 Hall introduced the sequential triple sampling procedure for point and interval estimation of the mean from a normal population with unknown variance. As its name implies, it involves up to three stages: a pilot stage, a main study stage and a fine tuning stage and it uses a stopping rule that mimics the fixed sample size rule, had the variance been known. The properties of this procedure are well understood in the case of exact normality of the underlying distribution. However, in practice the exact form of the underlying distribution is rarely known. In this paper we derive some general asymptotic results for the properties of the

procedure when the underlying distribution is unknown except that its first four moments are finite. More specifically, we consider underlying distributions that are in a neighbourhood of the normal, as are commonly used in classical robustness studies, to investigate the robustness of point and interval estimates derived from the sequential triple sampling procedure. Provided the underlying distribution is not too far from normal the asymptotic results are shown to agree well with estimated finite sample results. Some practical implications are also discussed.

The Sequential Feature of the Autoregressive Model

Christos P. Kitsos

Technological Educational Institute of Athens, Department of Mathematics, Athens, Greece
xkitsos@teiath.gr

The target of this paper is to discuss the well known, and particularly interesting from a theoretical point of view autoregressive model, which is also extensively applied. The estimation problems (from a linear problem we finally have a ratio estimate), the Fisher's information matrix (based on sequentially collected data points), and the problem on constructing confidence intervals (the variance is proportional to the inverse of the Fisher's information), are the main topics on were we are concentrated through a simulation study, as the model is also depended on the sample size.
Key Words: Fisher's information, sequential, autoregressive model

Evaluation of Robust Algorithms Sequence Designed to Eliminate Outliners from Cloud of 3D Points Representing 3D Human Foot Image

Samuel Kosolapov

Electrical and Electronic Engineering Department, ORT Braude Academic College of Engineering, Karmiel, Israel, ksamuel@braude.ac.il

In order to optimize operation of inexpensive non-laser 3D Structured Light Scanner adapted for fast Human Foot 3D Imaging, sequence of robust algorithms was designed, implemented and tested. During multi-frame image acquisition, sequence of Averaging and Median filters was used to reduce camera noise of different origin. A number of Robust Color Edge detectors used in the step of 3D calculations were evaluated and compared. The least amount of outliners was obtained in the algorithm evaluating Color Edge position by using scalar product in the RGB space after grey component elimination. Resulted cloud of 3D points was additionally processed by a sequence of logical operators and robust Median and Gaussian filters in order to eliminate 3D outliners. Processed cloud of 3D Points was overlapped onto 2D color image of the human foot (combined from a relevant frames of multi-frame image), thus creating true color presentation of the Human Foot. After visual inspection on the 2D monitor (for different view angles), created cloud of 3D points was converted to standard STL file and routed to 3D Printer for individual insole production. Accuracy and fitness of created insoles were evaluated and found accurate enough for chosen application.

On multivariate random probability measures

George Kouvaras[1] and George Kokolakis[2]

Department of Mathematics, National Technical University of Athens, Greece
[1] PhD student, gkouv@math.ntua.gr, [2] Professor of Statistics, kokolakis@math.ntua.gr

A Bayesian nonparametric model for the construction of multivariate random probability measures on a Polish space X is considered. The proposed model is built on a dependence structure among weighted gamma processes. Dependence is introduced by using a variant of Poisson/Gamma random field model. Results related to our approach are presented.
Keywords: Weighted gamma process, Poisson process, multivariate random probability measure.

Arithmetic Reduction of Intensity and Age models with bathtub failure intensity

Makram KRIT and Abdelwaheb REBAÏ

Institut des Études Technologiques de Gafsa, Cité des Jeunes, 2133 Gafsa, Tunisia
krit_makram@yahoo.fr

In this paper, we will study the estimation of maintenance e¢ ciency in Arithmetic Reduction of Intensity (ARI) and Arithmetic Reduction of Age (ARA) models with a memory m. These models have been proposed by Doyen (2005), the failure process is simply Non Homogeneous Poisson Process (NHPP). Our models are de.ned by reformulation of ARI and ARA ones using bathtub failure intensity. This form is presented like a superposition of two NHPP and Homogeneous Poisson one (HPP). Moreover, the particularity of this model allows taking account of system state improvement in time course. The maintenance effect is characterized by the change induced on the failure intensity before and after failure during degradation period. To simplify study, the asymptotic properties of failure process are derived. Then, the asymptotic normality of several maintenance e¢ ciency estimators can be proved in the case where the failure process without maintenance is known. Practically, the coverage rate of the asymptotic con.dence intervals issued from those estimators is studied.

Key Words: repairable systems reliability, bathtub failure intensity, imperfect repair, maintenance, Homogeneous Poisson Process, Non Homogeneous Poisson Process, estimation, likelihood.

Brown-Proschan imperfect repair model with bathtub failure intensity

Makrem KRIT and Abdelwaheb REBAI

Affiliation Institut Supérieur des Sciences Appliquées et de Technologie de Gafsa, Cité des jeunes, Gafsa, Tunisia, krit_makram@yahoo.fr

The aim of this paper is to study the estimation of maintenance efficiency in Brown-Proschan model. This model has been proposed by Brown and Proschan (1983), the failure process is simply a Non Homogeneous Poisson Process (NHPP). Our model is defined by BP reformulation one using bathtub failure intensity. This form of intensity is presented like superposition of two NHPP and Homogeneous Poisson one (HPP). Moreover, the particularity of this model allows to take account of system state improvement in time course. The characteristics of failure process and its ináuence on maintenance process are studied while basing on Monte-Carlo simulation. Finally, the main features of our model are derived: The likelihood function, thus parameter estimation and evaluation of maintenance efficiency are possible.
Key Words: repairable system, reliability, bathtub failure intensity, virtual age, imperfect maintenance, estimation, likelihood.

Flexible Quantile Regression for Survival Analysis

Milovan Krnjajić[1] and Athanasios Kottas[2]

[1] National University of Ireland, Galway, milovan.krnjajic@nuigalway.ie
[2] University of California at Santa Cruz, thanos@ams.ucsc.edu

Quantile regression is used to analyze the relationship between covariates and the higher or lower quantiles of the response distribution. Quantile regression can uncover important features which may be overlooked when using only standard mean regression. We develop Bayesian non-parametric models for quantile regression based on Dirichlet process mixtures (DPM) and dependent DPM-s, which allow the error distribution to change non-parametrically with covariates. We consider time-to-event data that include censored observations. Inference is based on implementation of a combination of posterior simulation methods for Dirichlet process mixtures. The performance of the proposed models is illustrated using simulated and real data sets.
Keywords: Bayesian non-parametrics, Dirichlet process mixture, censored data.

Application of Vector Optimization Methods in Learning Software Quality Evaluation Tasks

Eugenijus Kurilovas and Silvija Serikoviene

Vilnius University, Institute of Mathematics and Informatics; Vilnius Gediminas Technical University, Vilnius, Lithuania, eugenijus.kurilovas@itc.smm.lt
Vilnius University, Institute of Mathematics and Informatics, Vilnius, Lithuania, silvija.serikoviene@gmail.com

The main problem investigated in the paper is application and comparison of several vectorial optimization methods in learning software packages quality evaluation and optimization tasks. The paper reports on the research on the multiple criteria evaluation of quality and optimization of learning software packages (i.e., learning object repositories and virtual learning environment) using several vector optimization methods. Since solutions of vector optimization problems can be characterized and computed as solutions of appropriate scalar optimization problems, application of scalarization method using the experts' additive utility function is investigated in more detail. Several new elements make the given work distinct from all the other earlier works in the area.
Keywords: Vector optimization, Learning object repositories, Virtual learning environments, Evaluation of quality, Scalar optimization, Multiple criteria decision making

On Multivariate Extremality Measure

Henry Laniado, Rosa E. Lillo and Juan Romo

Department of Statistics, Universidad Carlos III de Madrid, Leganés, Spain
henry.laniado@uc3m.es ; hlaniado@est-econ.uc3m.es
Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain
rosaelvira.lillo@uc3m.es
Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain
juan.romo@uc3m.es

In general in the multivariate case there are not total orders. This problem has been partially solved with depth, which defines an order from the center outward [see, for example, Zuo and Serfling (2000) and Liu, Parelius and Sing (1999)].

We propose an alternative method for ordering multivariate data using a concept which we will call *Extremality*. This concept allows to establish the farness of $x \in R^n$ with respect to a cloud of points or to a distribution function, noting some relevant directions. Through this concept we will define a multivariate stochastic order. The natural properties are established for the new concept, asymptotic results are demonstrated, and some applications are shown.
Keywords: Extremality, statistical depth, multivariate stochastic order

Wentzel-Freidlin estimates in semi-group theory

Rémi Léandre

Institut de Mathématiques de Bourgogne, Université de Bourgogne, Dijon, France
Remi.leandre@u-bourgogne.fr

Abstract. We give the translation in semi-group theory of Wentzel-Freidlin estimates either for diffusion or jump processes. We give an application to Varadhan estimates for heat-kernel when we use a mixture of large deviation estimates and of the Malliavin Calculus of Bismut type translated by us in semi-group theory.
Keywords: Wentzel-Freidlin estimates, Varadhan estimates.

1 The case of a diffusion
In the first case we study the case of the small time behaviour of a diffusion. We translate in semi-group our proof of large deviation theory for diffusion which was valid for all the path space. Here this estimate is valid only for the semi-group. In particular, our proof is based upon the translation in semigroup theory of martingale exponential and of the *Itō* formula. By using a mixture between the Malliavin Calculus and large deviation theory, as it was pioneered by Bismut in his celebrated book "Large deviation and the Malliavin Calculus", we can translate in semi-group our proof of Varadhan estimate, upper bound, for a subelliptic heat-kernel which say when Wentzel-Freidlin estimate pass to heat-kernel.
2 The case of a Poisson process
In this case, we consider the behaviour of a jump process with more and more jumps which belong smaller and smaller. We show it is related to the theory of semi-classical expansion of Maslov and others people by looking the symbol of the operator. We translate in semi-group theory the proof of large deviation theorems of Wentzel-Freidlin which were valid for all the path space and now is valid only for the semi-group. We do a mixture between the Malliavin Calculus for jump process of Bismut type translated by ourself in semi-group theory and these Wentzel-Freidlin estimates in order to establish a logaritmic expansion of the involved heat-kernel.

References

1. Léandre, R., "Wentzel-Freidlin estimates in semi-group theory", In Control, Automation, Robotics and Vision, Yeng-Saih Choh and al eds, C.D.,I.E.E.E., 2233–2236, (2008).

2. Léandre, R., "Malliavin Calculus of Bismut type in semi-group theory", F.J.M.S. 30, 1–26 (2008).

3. Léandre, R., "Regularity of a degenerated convolution semi-group without to use the Poisson Process", To appear in "Non linear science and complexity II, M. Silva and al eds.

4. Léandre, R., "Malliavin Calculus of Bismut type without probability", In Fractional order systems", J.E.S.A., 715–733, (2008).

5. Léandre, R., "Wentzel-Freidlin estimates for jump process in semi-group theory: lower bound", In Int.Conf.Dif. Geom. Dynamical Systems", V. Balan edt, B.S.G. Proceedings 17, 107–113, (2010).

6. Léandre R., "Wentzel-Freidlin estimates for jump process in semi-group theory: upper bound", To appear Conf.Scientific. Computing, H. Arabnia edt, (2010).

7. Léandre, R., "Varadhan estimates for a degenerated convolution semi-group: upper bound". Preprint 596. Univ. Bourgogne (2010)

Polynomial structures in rank statistics distributions

Claude Lefèvre and Philippe Picard

Université Libre de Bruxelles, Département de Mathématique, Campus Plaine, C.P. 210, Bruxelles, Belgium, clefevre@ulb.ac.be

This paper deals with the classical problem of how to evaluate the joint rank statistics distributions for two independent i.i.d. samples from a common continuous distribution. It is pointed that these distributions rely on an underlying polynomial structure of negative binomial type. This property is exploited to obtain, in a systematic and unified way, various recursions, some well estabished, for computing the joint tail and rectangular probabilities of interest.
Key Words: Rank statistics; Kolmogorov-Smirnov two-sample tests; Generalized Appell and Abel-Gontcharoff polynomials; Negative binomial type polynomials; Lattice path enumeration

Models of Statistic Distributions of Nonparametric Goodness-of-fit Tests in Composite Hypotheses Testing in Case of Generalized Weibull Law

Boris Yu. Lemeshko, Stanislav B. Lemeshko and Kseniya A. Akushkina

Department of applied mathematics and computer science, Novosibirsk State Technical University, Novosibirsk, Russia, skyer@mail.ru

In this paper there are presented results (tables of percentage points and statistic distribution models) for the nonparametric goodness-of-fit tests in testing composite hypotheses using the maximum likelihood method for parameters estimation for Generalized Weibull Distribution law. Statistic distributions of the nonparametric goodness-of-fit tests are investigated by the methods of statistical simulation.
Keywords: Goodness-of-fit test, Composite hypotheses, Kolmogorov test, Cramer-Von Mises-Smirnov test, Anderson-Darling test, Generalized Weibull distribution.

Tests for homogeneity of variances under violation of normality assumption

Boris Yu. Lemeshko, Stanislav B. Lemeshko and Alice A. Gorbunova

Department of applied mathematics and computer science, Novosibirsk State Technical University, Novosibirsk, Russia, Lemeshko@fpm.ami.nstu.ru

The comparative analysis of power of classical variance homogeneity tests (Fisher's, Bartlett's, Cochran's, Hartley's and Levene's tests) is carried out. Distributions of tests statistics are investigated under violation of assumptions that samples belong to the normal law. Distributions and power of nonparametric tests of homogeneity of dispersion characteristics are researched (Ansari-Bradley, Mood's, Siegel-Tukey, Capon's, Klotz's tests). The comparative analysis of power of classical variance homogeneity tests with power of nonparametric tests is carried out. Tables of percentage points for Cochran's test are presented in case of the distributions which are different from normal.
Keywords: Test of homogeneity of variances, Fisher's test, Bartlett's test, Cochran's test, Hartley's test, Levene's test, Power of test, Nonparametric test, Ansari-Bradley test, Mood's test, Siegel-Tukey test, Capon test, Klotz's test.

Uniqueness and Extinction of Interacting Branching Collision Processes

Anyue Chen, Junping Li

Department of Mathematical Sciences, University of Liverpool, Liverpool L69 7ZL, UK,
achen@liv.ac.uk
School of Mathematical Sciences and computing technology, Central South University, China,
jpli@mail.csu.edu.cn

The uniqueness and extinction properties of the Interacting Branching Collision Process (IBCP), which consists of two strongly interacting components: an ordinary Markov branching process (MBP) and a collision branching process (CBP) will be addressed. We establish that there is a unique IBCP, and derive necessary and sufficient conditions for it to be non-explosive that are easily to be checked. Explicit expressions are obtained for the extinction probabilities for the regular, critical-explosive and super-explosive cases. The subtle sub-explosive case will also be addressed. The associated expected hitting times are also considered.
Keywords: Markov branching process; Collision branching process; Interacting branching collision process; Uniqueness; Extinction probability; Hitting times.

Monte Carlo conditioning on a sufficient statistic

Bo Henry Lindqvist and Gunnar Taraldsen

Department of Mathematical Sciences, Norwegian University of Science and Technology,
Trondheim, Norway, bo@math.ntnu.no
SINTEF Information and Communication Technology, Trondheim, Norway,
Gunnar.Taraldsen@sintef.no

Many statistical procedures are based on conditional distributions given sufficient statistics. The clue is that such distributions do not depend on unknown parameters. In certain cases one is able to derive the conditional distributions analytically. In general, however, it is impossible to arrive at tractable explicit expressions.The obvious solution is then to try Monte Carlo simulations. This talk is about such simulation approaches for computing conditional expectations given sufficient statistics, and also for sampling from the conditional distributions. Some applications will be given, as well as a brief discussion of connections to Bayesian and fiducial statistics.
Keywords: Sufficiency, Conditional distribution, Monte Carlo simulation.

References
1. Lindqvist, B.H. and Taraldsen, G., Monte Carlo conditioning on a sufficient statistic, *Biometrika,* **92**, 451-464 (2005).
2. Lockhart,R., O'Reilly, F.J. and Stephens, M.A., Use of the Gibbs sampler to obtain conditional tests, with applications, *Biometrika,* **94**, 992-998 (2007).

Personalization of Text Information Retrieval with Bayesian Networks and Evolutionary Algorithms

Piotr Lipinski

Institute of Computer Science, University of Wroclaw, Wroclaw, Poland, lipinski@ii.uni.wroc.pl

Information retrieval (IR) focuses on searching through large databases of unstructural information and finding documents relevant to queries sent by users. Queries are usually imprecise and often do not reflect exact user intentions. In some cases, users send the same query many times or want to continuously monitor databases (changing frequently - each time

when new documents appear), which enables to collect some user feedback, analyze it to detect hidden user intentions and tune further searching.

Each document is described by two vectors: $\mathbf{c} \in C_1 \times C_2 \times \ldots \times C_K$, related to meta data (such as authors, languages, publishers, keywords), where C_i denote domains of such meta data components, and $\mathbf{d} \in \mathbf{R}^L$, related to content-based data, whose components correspond to significance of some pre-selected terms in the document.

In the classic approach, each query Q consists of two parts: a constraint on meta data and a list of significant terms, which, in a way, defines a subset of $\Omega = C_1 \times C_2 \times \ldots \times C_K \times R^L$.

In the approach proposed, each query Q has assigned a probability distribution PQ over - defined by a Bayesian Network (BN). At the beginning, P_Q is concentrated on $Q \in \Omega$, with uniform probability over Q, which corresponds to the original query Q. Using the probability distribution, the IR system draws a random sample of query instances and finds documents relevant to the sample. After getting user feedback, the IR system updates the probability distribution to better fit user evaluations by solving an optimization problem using evolutionary algorithms (EAs) estimating a new probability distribution and constructing a new BN. IR continues to search through databases with the new probability distribution and all the process repeats until the user stops monitoring databases.

Results of some experiments performed on real-life data prove that such a system is able to personalize IR, adapt to hidden user intentions and increase accuracy of results. EAs applied in the IR system, based on the Bayesian Optimization Algorithm (BOA), are able to find efficient solutions for the optimization problem and construct a BN precisely describing user feedback.

Low-Rank Data Format for Uncertainty Quantification

Alexander Litvinenko and Hermann G. Matthies

Institute of Scientific Computing, Technische Universität Braunschweig, Germany
wire@tu-bs.de

We research how uncertainties in the input data (parameters, coefficients, right-hand sides, boundary conditions, computational geometry) spread of in the solution. Since all realisations of random fields are too much information, we demonstrate an algorithm of their low-rank approximation. This algorithm is based on singular value decomposition and has linear complexity. This low-rank approximation allows us to compute main statistics such as the mean value, variance, exceedance probability with a linear complexity and with drastically reduced memory requirements.

Keywords: uncertainty quantification, stochastic elliptic partial differential equations, Karhunen-Loève expansion, QR-algorithm, sparse data format, low-rank data format.

The Homogeneous Markov System (HMS) as an elastic medium. The three-dimensional case

[1]Maaita J. O., [2]Tsaklidis G. and [1]Meletlidou E.

[1]Department of Physics, [2]Department of Mathematics, Aristotle University of Thessaloniki, Greece

Every attainable structure of the continuous-time Homogeneous Markov System (HMS) with fixed size and state space $S=\{1,2,\ldots,n\}$ is considered as a particle of $\mathbf{R}^n$ and consequently the motion of the structure corresponds to the motion of the particle. Under the assumption that "the motion of every particle-structure at every time point is due to its interaction with its surroundings", $\mathbf{R}^n$ becomes a continuum.

Then the evolution of the set of the attainable structures corresponds to the motion of the continuum. For the case of a three-state HMS it is stated that the concept of the two-dimensional isotropic elasticity can further interpret the HMS's evolution.
Key Words: Markov Systems, Population dynamics, Continuous medium

A DECADE OF COVARIANCE MODELLING AND BEYOND

Gilbert MacKenzie[1,2] and Jing Xu[2]

[1]ENSAI, Rennes, France, gilbert.mackenzie@ul.ie
[2]Centre for Biostatistics, University of Limerick, Ireland, jing.xu@ul.ie

The conventional approach to modelling longitudinal data places considerable emphasis on estimation of the mean structure and much less on the covariance structure, between repeated measurements on the same subject. Often, the covariance structure is thought to be a "nuisance parameter" or at least not to be of primary scientific interest" and consequently little effort is expended on modelling its structure. In particular, the idea that in longitudinal RCTs intervention might affect the covariance structure rather than, or as well as, the mean rarely intrudes.

A decade on, we shall argue that these ideas are rather passé and that from an inferential standpoint the problem is symmetrical in both parameters μ and Σ. Throughout, we will distinguish carefully between joint estimation, which is now relatively routine, and joint model selection, which is not.

At first sight the task of estimating the structure of Σ from the data, rather than from a pre-specified menu, may seem daunting, whence the idea of searching the entire covariance model space, C for Σ, may seem prohibitive. Thus, the final demand that we conduct a simultaneous search of the Cartesian product of the mean-covariance model space, $M \times C$ may seem impossible. However, below, we shall accomplish all three tasks elegantly for a particular, but very general, class of covariance structures, C^*, defined below.

The technique is based on a modified Cholesky decomposition of the usual marginal covariance matrix $\Sigma(t;\theta)$ where t represents time and θ is a low-dimensional vector of parameters describing dependence on time. The decomposition leads to a reparametrization, $\Sigma(t;\xi,\phi)$ in which the new parameters have an obvious statistical interpretation in terms of the natural logarithms of the innovation variances, ξ, and autoregressive coefficients ϕ. These unconstrained parameters are modelled, parsimoniously, as different polynomial functions of time.

In this talk we trace the history of the development of joint mean-covariance modelling over the last decade, from Pourahmadi's seminal paper in 1999 to recent times, discuss recent advances in research in the paradigm and remark on the potential impact on longitudinal trial design and analysis.

Meta/Exact hybridization to improve intensification and diversification in a multi-objective context

Samir Mahdi, Salim Chikhi and Mohamed Batouche

Univercite Mentouri Labo Misc, Contantine, Algeria, s.mahdi_dz@yahoo.fr

Technical impossibility to solve exactly NP-hard combinatorial optimization problems for large instances requires the use of heuristics and especially metaheuristics. Nevertheless, the exact methods can be useful, when subproblems can be extracted from the whole problem. Indeed, their resolution contributes in the global solution search, by combining exact resolution of subproblems and heuristic resolution of the complete problem. This approach is generally efficient, because it combines the advantages of two

very different methods. An efficient method must establish a good compromise between the convergence toward the Pareto frontier (intensification) and the distribution of the solutions along the Pareto frontier (diversification). In this paper we propose to hybridize the metaheuristic MA|PM (Memetic Algorithm with Population Management, adapted to multi-objective problem) with Branch & Bound algorithm t! o solve combinatorial multi-objective optimization problems based on the concept of Pareto optimum. In the metaheuristic we have combined methods with the capacity of intensification and methods which have faculties of diversification, in order to obtain a good compromise between convergence towards the Pareto frontier and the distribution of the solutions along the Pareto frontier. Our idea is to add there, an exact method, which has an absolute research power, in order to improve the wished compromise between the intensification and the diversification. We have realized experiments on well-known benchmarks in the literature of the multi-objective knapsack problem. The results obtained show the good behavior of our method, and the comparison between (MA|PM) and our method shows the effectiveness of Meta/Exact hybridization.

Key Words: Meta/Exact hybridization, branch & bound, memetic algorithm, population management, diversification, intensification.

The exact control limits of an EWMA control chart with estimated parameters

Petros E. Maravelakis

University of the Aegean, Department of Statistics and Actuarial-Financial Mathematics, Karlovasi, Samos, Greece

A usual assumption in the development of control charts is that the process parameters are known. However, in practice they are usually estimated from a historical (Phase I) data set. It has been shown that the performance of control charts is greatly affected when the parameters are estimated by a few number of samples. In this paper we compute the exact control limits of an EWMA chart when the parameters are estimated so that the performance of the chart is not affected whether we estimate the parameters or not.

Tail behaviour of β-ARCH processes

László Márkus, Péter Elek

Dept. Probability Theory and Statistics, **Eötvös Loránd** University, Budapest, Hungary
markus@cs.elte.hu

It is now common knowledge that the simple quadratic ARCH-process has regularly varying tail even when generated by a normally distributed noise, and the tail behaviour is well-understood under more general conditions, as well. Much less studied is the case of β-ARCH-type processes, i.e. when the conditional variance is a 2β-power function, $0 < \beta < 1$. This class is popular in modelling such financial and environmental phenomena, when almost all "stylised facts" for GARCH models are observable, but - in contrast to the quadratic case - light tails in the noise do not generate heavy tails. (Unlike the stock-exchange, Mother Nature does not panic). Being a little more general and allowing for asymmetry, we consider threshold β-ARCH models, driven by Weibull-type noises. Such a broad class of noises include e.g. the Gaussian and the double exponential ones as special cases. We show that the generated process has approximately Weibull like tail, too, albeit with different exponent: $1-\beta$ times that of the noise. We cannot, however, prove exact asymptotic tail behaviour only bound the tail from both side by Weibull distributions of the same exponent but slightly different constants. The proof is based on taking an appropriate auxiliary sequence and then apply the general result of Klueppelberg and Lindner for the tail of infinite MA sequences.

Classification and Combining Models

Anabela Marques, Ana Sousa Ferreira and Margarida Cardoso

ESTBarreiro, Setúbal Polytechnic, Portugal, CEAUL, anabela.marques@estbarreiro.ips.pt
LEAD, Faculty of Psychology, University of Lisbon, Portugal, CEAUL, asferreira@fp.ul.pt
UNIDE, Dep. of Quantitative Methods of ISCTE - Lisbon University Institute, Portugal,
margarida.cardoso@iscte.pt

In the context of Discrete Discriminant Analysis (DDA) the idea of combining models is present in a growing number of papers aiming to obtain more robust and more stable models. This seems to be a promising approach since it is known that different DDA models perform differently on different subjects. Furthermore, the idea of combining models is particularly relevant when the groups are not well separated, which often occurs in practice.
Recently, we proposed a new DDA approach which is based on a linear combination of the First-order Independence Model (FOIM) and the Dependence Trees Model (DTM). In the present work we apply this new approach to classify consumers of a Portuguese cultural institution. We specifically focus on the performance of alternative models' combinations assessing the error rate and the Huberty index in a test sample.
We use the R software for the algorithms' implementation and evaluation.
Keywords: Combining models, Dependence Trees model, Discrete Discriminant Analysis, First Order Independence model.

Geostatistical predictions based on block data

Shigeru Mase[1], Kouji Takahashi, and Nobuyuki Mochida

Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan, mase@is.titech.ac.jp

Block-to-block and block-to-point kriging predictions based on block data are proposed. Blocks may be regular (mesh data) or of more general shapes. Under the assumptions of second-order stationarity and isotropicity, we show how to lessen the number of calculations of relevant block-to-block covariances. As illustrations, a mesh data of population and a simulated block data on convex polygons are analyzed.
Keywords: Geostatistics, Kriging predictor, Block data, Block-to-point prediction, Mesh data, Change of scale problem, Gaussian Random Field.

Drag-Free, Attitude and Formation Control for Monitoring the Earth Gravity Field

Luca Massotti[1], Andres Molano[2], Javier Jimenez[2] and Enrico Canuto[2]

[1] ESA ESTEC, Earth Observation - Future Missions Division (EOP-SFP), 2200 AG, Noordwijk ZH, The Netherlands, luca.massotti@esa.int
[2] Politecnico di Torino, Dipartimento di Automatica e Informatica, Corso Duca degli, Abruzzi 24, 10129 Torino, Italy, enrico.canuto@polito.it

Following the successful launch in Spring 2009 of the GOCE satellite (Gravity field and steady-state Ocean Circulation Explorer), a more ambitious mission consisting in a satellite formation of two satellite separated by (a minimum) 10 km distance, is under study at the European Space Agency, aiming at monitoring the Earth's gravity field fluctuations, during a (at least) 6-year mission. Since GOCE is the first flying drag-free satellite, the envisaged formation might be the first drag-free formation, posing a suite of challenging technology and control problems under study and solution. The paper concentrates on a triad of control problems to be solved and coordinated (formation, drag-free and attitude), all of them being constrained by a long-life low-Earth-orbit mission, imposing low propellant mass, scarce electric propulsion throttability and limited electric power. Driving requirements are presented

and discussed, showing how they can be met through Embedded Model Control design. Finally, realistic simulated results are included.

Keywords: Satellite formation, control, drag-free, attitude, low-Earth-orbit, gravity monitoring,

Stochastic Modelling and Comparative Analysis for the Demographic Data of Scandinavian Countries

George Matalliotakis[1], Christos H. Skiadas[2], Vardoulaki Maria[3]

Technical University of Crete, Chania, Crete, Greece
Emails: [1] matalliotakis@hotmail.com , [2] skiadas@ermes.tuc.gr , [3] mv_math@yahoo.gr

In this article we report some demographic characteristics of Scandinavian countries focusing on the different mortality features of both sexes between male and female population during the last century. From the data analysis it is observed the gradual decrease of "children's mortality" and the increase of the mean life span. Findings also are emphasizing the changes in the death rate and how it can be affected by events (war, disease).

A modelling approach to Life Table Data sets is proposed. The method is based on a stochastic methodology and the derived first exit time probability density function. The Health State Function $H(t)$ of the population is modeled as the mean value of the health states of the individuals and is assumed to be at a specific level at the time of birth and then changes and gradually decreases to a zero level at the time of death.

A new, simple model based on a stochastic theory introduced by Janssen and Skiadas (1995), and modified recently by Skiadas and Skiadas (2010), is presented and called S – model.

$$g(t) = \frac{|H_t - tH_t'|}{t} \, p(t) = \frac{|H_t - tH_t'|}{\sigma\sqrt{2\pi t^3}} \, e^{-\frac{(H_t)^2}{2\sigma^2 t}}$$

The first exit time density $g(t)$ of the stochastic model proposed ($p(t)$ is the probability density function of the stochastic process) for the Health State Function $H(t)$ is compared to the three parameter Weibull and Gompertz models. The three models are applied to the Life Table Data for males and females in Scandinavian countries. The results will be presented and analyzed. The conclusions indicate that the proposed model fits better to the data than the Weibull and Gompertz models whereas has a significant explanatory ability. The methodology od the model building and the model proposed could be used in several cases in population studies in biology, ecology and in other fields.

1. Janssen, J. and Skiadas, C. H., Dynamic Modelling of Life-Table Data. Applied Stochastic Models and Data Analysis, vol. 11, No 1, 35-49, 1995.
2. Skiadas, C. and Skiadas, C. H., Development, Simulation and Application of First Exit Time Densities to Life Table Data, Communications in Statistics - Theory and Methods, Volume 39, Issue 3 January 2010, pages 444 - 451.

A Dynamic Model of Life Table Data Analysis for the population of various countries

George Matalliotakis[1], Christos H. Skiadas[2], Vardoulaki Maria[3]

Technical University of Crete, Chania, Crete, Greece
Emails: [1] matalliotakis@hotmail.com , [2] skiadas@ermes.tuc.gr , [3] mv_math@yahoo.gr

In this poster we performed a methodology of stochastic processes of life table data. An analysis of the mortality curves follows, and comments about their shape, their course during time, and differences between males and females are studied. We refer to indicators concerning the increase in average life, the possible change in the number of births, deaths and immigrants.

The method is based on a stochastic methodology. A modeling approach to Life Table Data sets is proposed. A new, simple model, introduced by Skiadas and Skiadas (2010) is presented and applied to data sets. The health state function is defined and estimated for the population too. We use the life table data to apply the specific methodology of the first exit time of a stochastic process and we compare the resulting probability density function to the related forms of the Weibull and Gompertz. The results the three models are displayed in graphs and compared.

The study shows that the S-model gives lower errors, and describes the life table data distribution with greater accuracy, than the Weibull and Gompertz models. The S-model seems to be more appealing for Life Table Data sets analysis due to the explanatory ability inherent.

Skiadas, C. and Skiadas, C. H., Development, Simulation and Application of First Exit Time Densities to Life Table Data, Communications in Statistics - Theory and Methods, Volume 39, Issue 3 January 2010, pages 444 - 451.

Decision-Making Criteria for Electricity Customer Satisfaction and Financial Success

Rabah Medjoudj , Djamil Aissani and Klaus Dieter Haim

University of Bejaia, Lamos laboratory, route targua ouzemour, Bejaia, Algeria
medjoudj@yahoo.fr

The aim of this paper is to apply two different multi-criteria methods for decision making in electrical distribution system operation and planning. The first method is based on the analytical hierarchy process (AHP) and the second is based on game theory. In both cases, the goal to attempt is customer satisfaction and company financial success. The criteria are: cost, reliability, availability, maintainability and quality. As for alternatives, they correspond to technical and organizational measures to develop and to be translated through automation and restructuring of the network. From a practical standpoint, the obtained results provide decision makers with a range of choices enabling them to target a well-defined goal and take appropriate action according to the company means. Regarding the method inspired by game theory often applied to competitive or liberal environment, in this paper, it was shown tha! t it can be applied in a monopoly one taking into account the customer's reaction against quality of service level, and scales well with the actual financial crisis.
Key Words: Electrical distribution system, AHP model, game theory, reliability indices.

Weibull proportional hazards model to modeling the impact of somatic cell on the length of productive life of Tunisian dairy cattle

Naceur M'HAMDI[1a], Rafik ALOULOU[1], Mahdi BOUALLEGUE[2], Satinder K. BRAR[3], and Mohamed BEN HAMOUDA[4]

[1]Department of Animal Sciences- Higher Institute of Agronomy of Chott-Mariem, BP47, 4042. Sousse, Tunisia
[2] Faculty of Mathematic Physic and Natural Sciences of Tunis, Tunisia.
[3] National Institute for Scientific Research, Quebec, Canada.
[4]Institution of Agricultural Research and Higher Education, (IRESA), Ministry of Agriculture and Hydraulic Resources. Tunisie
[a]Corresponding address: Department of Animal Sciences- Higher Institute of Agronomy of Chott-Mariem, BP47, 4042. Sousse, Tunisia
Tel: (216) 73 348 544- Fax: (216) 73 327 591, naceur_mhamdi@yahoo.fr

The ability of a dairy cow to resist involuntary culling is of high economic importance to the dairy farmer. Genetic improvement of functional longevity can lead to lower replacement costs, decreased veterinary costs, more opportunities for voluntary culling, and a higher proportion of cow's productivity at a mature level. The relationship between somatic cell

scores (SCS) and longevity for Tunisian Holstein–Friesian dairy cows was assessed using Weibull proportional hazards and Cox Frailty models. Data from the first three lactations of 7120 Tunisian Holstein cows having their first calving between 1996 and 2003 was used to perform the calculations. The results showed that a cow with a test-day in the highest class of phenotypic level for SCS showed more than three times greater risk of being culled when compared to risk for the class with the lowest SCS level. The genetic correlation between the risk of being culled and SCS was estimated to be 0.37. Udder health problems were the reason of culling for 23% of the culled cows in this study. For each unit increase in the log-transformed SCC, the culling hazard increased by 26% Hazard ratio (HR) = 1.26]. Culling decision of farmers of Tunisian Holstein–Friesian dairy cows and SCS is thus genetically associated to the functional longevity of cows.

Keywords: Tunisian Holstein–Friesian dairy cows; risk; somatic cell scores; longevity

Variation factors of Functional LongevityTunisian Holstein- Friesian population cows

Naceur M'HAMDI[1a], Rafik ALOULOU[1], Mahdi BOUALLEGUE[2], Satinder K. BRAR[3], and Mohamed BEN HAMOUDA[4]

[1]Department of Animal Sciences- Higher Institute of Agronomy of Chott-Mariem, BP47, 4042. Sousse, Tunisia
[2] Faculty of Mathematic Physic and Natural Sciences of Tunis, Tunisia.
[3] National Institute for Scientific Research, Quebec, Canada.
[4]Institution of Agricultural Research and Higher Education, (IRESA), Ministry of Agriculture and Hydraulic Resources. Tunisie
[a]Corresponding address: Department of Animal Sciences- Higher Institute of Agronomy of Chott-Mariem, BP47, 4042. Sousse, Tunisia
Tel: (216) 73 348 544- Fax: (216) 73 327 591, naceur_mhamdi@yahoo.fr

Description of genetic evaluation for longevity in Tunisian Holstein-Friesian population is presented. Functional longevity (FL), measured as length of productive life (LPL) adjusted by production level, is an important trait to measure the global functional aptitude of a cow, particularly when direct measures of the functional traits (health and reproductive performance) are not available. Using a proportional hazard mixed model stratified by herd, the influences of age at first calving, herd- year- season, lactation number- stage of lactation, milk production within the herd and year, annual change in herd size and random genetic effect of sire on length of productive life were studied. Two models were explored: the first one included time-dependent variables taking the current value of the covariate at each calving date, while the second one also comprised time-dependent interactions between the values of the covariate of the current calving and its mean value during the last three previous calving. Besides a random time-dependent effect of herd–year–season, the model included fixed effects of parity and stage of lactation, within herd and parity production level (all time-dependent), and age at first calving (time-independent). All factors had a significant effect on LPL at a level of p <0.001. Milk yield was by far the most important factor influencing LPL and the hazard increased as the milk production decreased. Within the first three lactations, the relative culling risk increased stepwise with lactation stage. The highest risk of culling was found for cows at the beginning and at the end of the first lactation and at the end of any other lactation. The risk of culling decreased with parity.

The Flexible Dirichlet family: some inferential issues

S. Migliorati, G. S. Monti, and A. Ongaro

Department of Statistics, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy. sonia.migliorati@unimib.it

In many problems data consist of vectors of proportions, such as chemical constituents of a substance, and are, therefore, subject to a unit sum constraint. This type of data, called

compositional, arise naturally in a great variety of disciplines such as archeology, biology, economics, environmetrics, psephology, medicine, psychology, etc..

The most well known distribution for compositional data is the Dirichlet which possesses several good statistical and mathematical properties, such as closure under amalgamation and subcomposition, as well as easiness of parameter interpretation. However it is only suitable for modeling data exhibiting the maximum degree of independence compatible with compositions.

The Flexible Dirichlet distribution (Ongaro, A., Migliorati, S., Monti, G.S., *A new distribution on the simplex containing the Dirichlet family*, Proceedings of the CODAWORK08, Girona, Spain, 2008) is a generalization of the Dirichlet which preserves some of its good mathematical properties and, at the same time, exhibits a complex dependence structure which allows various forms of dependence relevant for compositional data (e.g. non-neutrality, subcompositional dependence and subcompositional non-invariance), independence cases being identified by suitable parameter configurations.

Here we shall investigate some inferential aspects of the Flexible Dirichlet, which require particular attention due to the mixture structure and the special parameter configuration of the distribution. More precisely, we shall develop suitable likelihood-based testing procedures to assess the presence of (in)dependence relations (such as neutrality) of particular impact in applications. Their performances will be evaluated by means of Monte Carlo experiments. Applications to real data will also be presented.

Keywords: Generalizations of Dirichlet distribution, Finite mixture, Compositional data, Neutrality, E-M algorithm, Likelihood.

COMPOUND BINOMIAL RISK MODEL

Leda D.Minkova

Faculty of Mathematics and Informatics, Sofia University "St. Kl.Ohridski" Bulgaria
leda@fmi.uni-sofia.bg

In this note the counting process in the insurance risk model is a compound Binomial process. The particular case of geometric compounding distribution is analyzed. The counting process is called Inflated-parameter binomial process (I - Binomial process). Some basic properties are given. The corresponding risk model is called I- Binomial risk model. The joint probability distribution of the time to ruin and the deficit after ruin occurs is studied. The case of exponentially distributed claims is given.

Families of Distributions Arising from Distributions of Record Statistics*

S. M. T. K. Mirmostafaee[1] and Jafar Ahmadi[2]

[1] Ferdowsi University of Mashhad, P. O. Box 91775-1159, Mashhad, Iran
(e-mail: ta¡mi182@stu-mail.um.ac.ir)
[2] (e-mail: ahmadi-j@um.ac.ir)

This paper proposed a new general family of continuous distributions motivated by the distributions of record statistics. Its distributional properties including the distribution function, moments, symmetry and modality are studied.

One special case, when F is the exponential distribution is considered and at the end two real data sets are used for fitting the suitability of our proposed model in the special case.

Keywords: Order statistics, Record statistics, Gamma distribution.

The Network Design Problem with Stochastic Game

Hiroaki Mohri

Waseda University, Nishi-Waseda 1-6-1, Shinjuku-Ku, Tokyo, 169-8050 JAPAN
mohri@waseda.jp

In this paper, we present an analysis for the "Network Design Problem with Stochastic Game". The Network Design Problem (sometimes, called "Network Synthesis Problem") is a minimum multi commodity flow problem with fixed arc cost. It is applied mainly for computer or telecommunication networks. But we should expand these networks on demand, i.e. we have to decide "when" and "how to" expand them. At this point, we apply "Stochastic Game" to this problem. On expanding networks, we decompose a potential underlying graph. Regarding each decomposed graph as a player, we suppose a hidden Markov Game. This method might be applied for transportation, biological and other networks.

Nonparametric Inference in the Class of Controlled Two-sex Branching Models

Manuel Molina, Manuel Mota, Alfonso Ramos

Department of Mathematics, University of Extremadura, Spain
mmolina@unex.es, mota@unex.es, aramo@unex.es

Inside the general context of stochastic modeling, the branching process theory provides stochastic models to describe the probabilistic evolution of systems whose components after certain life period reproduce and die. Nowadays, it is an active research area of both theoretical interest and applicability to several fields, see for example Haccou *et al.* (2005). In particular, with the purpose to model the probabilistic evolution of populations where females and males coexist and form couples (female-male), some classes of two-sex branching models have been investigated. However, the range of models studied is not large enough in order to get an optimum modeling in many two-sex populations where a control on the number of couples in the population is required. It can be stated that significant efforts have been made regarding random control branching models with asexual reproduction. Now similar efforts should be made to develop models with a random control where reproduction is bisexual. In an attempt to contribute some solution to this problem, recently Molina *et al*. (2010) have introduced the class of two-sex branching models with random control on the number of progenitor couples. In such a paper, some probabilistic results have been established and, assuming offspring distribution belonging to the bivariate power series family, estimators for the mean vector and the covariance matrix of the offspring distribution have been proposed. In this work, we continue the research on this class of two-sex models. In fact, by considering that no assumptions are made about the functional form of the underlying offspring probability distribution, we obtain Bayes estimators for the offspring probability law and for its main moments. Also, we determine the corresponding 95% highest posterior density credibility sets. By way of illustration, we present some simulated examples where we check the accuracy of both the estimates and their corresponding 95% highest posterior density credibility sets.

Keywords: Branching models, Two-sex models, Controlled models, Nonparametric inference, Bayesian inference.

References
1.Haccou, P., Jagers, P., Vatutin, V. (2005). *Branching Processes: Variation, Growth, and Extinction of Populations.* Cambridge University Press.
2.Molina, M., Mota, M. And Ramos, A. Two-sex branching models with random control on the number of progenitor couples. *Methodology & Computing in Applied probability (to appear).*

Extinction Probability in the Class of Two-Sex Branching Models with Offspring and Mating Depending on the Number of Couples

Manuel Molina[1], Yongsheng Xing[2], Shixia Ma[3]

[1] Department of Mathematics, University of Extremadura, Badajoz, Spain, mmolina@unex.es
[2] College of Mathematics, Shandong Institute of Technology, Yantai, China, xingys@nankai.edu.cn
[3] School of Sciences, Hebei University of Technology, Tianjin, China, maxhixia1@163.com

Mathematical models in which individuals may be replaced by others of a similar or different type are now being developed. It is an active research area of both theoretical interest and applicability to such fields as biology, demography, ecology, epidemiology, genetics, population dynamics, and so on. One may cite, for example, the monographs by Haccou *et al*. (2005) and Kimmel and Axelrod (2002) which include practical applications to cell kinetics, cell biology, chemotherapy, human evolution, and molecular biology. In particular, it is our purpose to model the probabilistic behavior of populations where females and males coexist and form couples (female-male). In this work, we continue the research concerning the class of two-sex models with offspring and mating depending on the number of couples in the population introduced in Molina *et al.* (2008). In addition its theoretical interest, such a class of stochastic models has also practical implications, especially in population dynamics. We consider the extinction probability problem. We establish some necessary and sufficient conditions for the almost sure extinction of the two-sex branching model. Also, we investigate its stochastic comparison with others classes of two-sex models studied in branching process literature. As illustration, we provide several simulated examples.
Keywords: Branching models, Two-sex models, Extinction probability.

References
1. Haccou, P., Jagers, P., Vatutin, V. (2005). *Branching Processes: Variation, Growth, and Extinction of Populations.* Cambridge University Press.
2. Kimmel, M. And Axelrod D.E. (2002). *Branching Processes in Biology.* Springer.
3. Molina, M., Jacob, C. and Ramos, A. Bisexual branching processes with offspring and mating depending on the number of couples in the population. *Test, 17, 245-281, (2008).*

Statistical Approach to Transport of Modulation Instabilities in Optic Fibers

Rasulova Mukhayo

MIMOS BHD (Malaysian Institute of Microelectronics)
MIMOS BHD., Technology Park Malaysia, Bukit Jalil, 57000, Kuala Lumpur, Malaysia
rasulova@live.com

In the paper we prove equality of wave equation for Modulation Instability and Vlasov equation. The quantum and quasi-quantum analogues of wave equation for Modulation Instability is defined. On the basis of BBGKY hierarchy of quantum kinetic equations the kinetic equations for any number Modulation Instabilities are defined and the method solution for these equations are proposed
Key Words: Modular Instability, Vlasov equation, von-Noymann equation, BBGKY hierarchy

Macro effects of global economic crisis on trade, economic growth and Poverty in India

KN Murty

Department of Economics, School of Social Sciences, University of Hyderabad, Gachibowli- 500 046, Hyderabad, India. Email: knmurty@yahoo.com.

This paper attempts to use an aggregative, structural, macro-econometric model for India that has been under development by the author and his associates in recent years. The model emphasizes the inter-relationships between internal and external balances and also the relation between money, output, prices and balance of payments. It also tries to link economic growth with poverty reduction in a very simplistic way. In the estimated model, investment and output are disaggregated into four sectors, viz., (i) agriculture including forestry & fishing, (ii) manufacturing including mining, (iii) infrastructure, which includes power, transport, communication and construction and (iv) services sector, covering all other activities in the economy. Annual time series data for 1978-'79 to 2002-'03 and 3SLS method with simultaneously iterative weighing matrix and coefficients are used, which gave more efficient estimates of the parameters and better goodness-of-fit statistics. The estimated model has 74 endogenous variables of which 35 are equations and 39 identities. It also has 39 exogenous variables including 8 dummy variables. The model is validated for its in-sample and out-of-sample (2003-'06) forecasting ability. Though the model did fairly well for in-sample forecasting for many variables, the good performance is limited to fewer variables in the case of out-of-sample forecasting. This study further attempts counter factual stochastic simulations relating to global economic crisis, trade liberalization and stepping-up public investment in infrastructure as a way of overcoming the global economic crisis in India.

Key words: Macro econometric model, stochastic simulation, and global economic crisis.
JEL Codes: C15, C30, C530 and E27.

A new concept in possibility of equality to create innovative constraints in fuzzy linear regression

R. Nadimi , S. F. Ghaderi

Department of Industrial Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran, r.nadimi@yahoo.com

Possibility of equality between two or more fuzzy numbers is a popular method to consider their degree of fitness. Possibility of equality may be applied to establish the constraints of fuzzy linear regression in which conjunction problem is under consideration. In this study, a new concept of the possibility of equality, that creates new restrictions, will be introduced and applied in fuzzy regression model, and then a more precise method will be represented to calculate the amount of error. To compare the performance of the proposed approach with the other methods, numerical examples are given. Total amount of error is calculated to confirm the efficiency of the proposed approach.
Keywords: Fuzzy linear regression; Possibility of equality; Fuzzy number.

Factor Analysis (FA) as ranking and an Efficient Data Reducing approach for decision making units

Reza Nadimi[*], Hamed Shakouri G. [*], Reza Omidi[+]

[*]Department of Industrial Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran.
[+]Department of Social Policy, Faculty of Social Science, University of Tehran, Tehran, Iran.
h.shakouri@gmail.com

This article compares two techniques: Data Envelopment Analysis (DEA) and Factor Analysis (FA) to aggregate multiple inputs and outputs in the evaluation of decision making units (DMU). Data envelopment analysis (DEA), a popular linear programming technique, is useful to rate comparatively operational efficiency of DMUs based on their deterministic or stochastic input–output data. Factor analysis techniques, such as Principal Components Analysis, have been proposed as data reduction and classification technique, which can be applied to evaluate of decision making units (DMUs). FA, as a multivariate statistical method, combines new multiple measures defined by inputs/outputs. Nonparametric statistical tests are employed to validate the consistency between the ranking obtained from DEA and FA. Also, the results have been compared with PCA approach. Results of numerical reveal that new approach shows a consistency in ranking with DEA.

Keywords: Decision Making; Data Envelopment Analysis; Factor Analysis, Principal Component Analysis.

MCMC Estimation and Imputation in Autologistic Model with Auxiliary Variables

Wojciech Niemiro and Marta Zalewska

Nicolaus Copernicus University, Toruń, Poland and University of Warsaw, Poland
wniemiro@gmail.com
Medical University of Warsaw, Poland, zalewska.marta@gmail.com

Autologistic probability distribution (Bessag, 1974) is a flexible model for dependent binary data, especially in applicatons to epidemiological surveys. Estimation of parameters by maximum pseudolikelihood makes use of the generalized linear models (GLM) methodology. Geyer and Thompson (1992) proposed to compute maximum likelihood estimates (MLE) using Markov chain Monte Carlo (MCMC). Gibbs Sampler (GS) is particularly suitable because of the simple form of full conditionals.

In this paper we present a method of estimation which is an MCMC approximation to one-step-maximum likelihood. We show that the extra variability introduced by simulating Monte Carlo samples can be used to advantage. Instead of MLE we compute an asymptotic approximation of a posterior distribution in a Bayesian version of autologistic model. We show that the estimates obtained in this way can be incorporated into an algorithm for imputing missing data. We also consider a model with auxiliary variables, both cathegorical and numeric.

We report results of analyses of epidemiological data, collected in a big survey ECAP (Prevalence of allergic diseases in Poland 2006-2008). We fit autologistic model and impute missing values of binary responses. Some simulated experiments are also included, to assess the performance of our method. Our algorithms are implemented as R functions.

Keywords: Markov chain Monte Carlo, Missing values, Logistic Regression, Importance Sampling, Bayesian Statistics, Asymptotic Approximation, Generalized Linear Models, R.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J.R. Statist. Soc. B*, 36, 192--236.

Geyer, C.J. and Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J.R. Statist. Soc. B*, 54, 3, 657--699.

Strategies and methodologies of Experimental Design in the online environment

Teresa Oliveira and Amilcar Oliveira

CEAUL and DCeT - Universidade Aberta, Lisboa, Portugal

Experimental Design is a branch of research in many areas, very varied structures and with many answers not yet known, being one of the most fascinating fields of research in Statistics. It has underlying ideas as important and in vogue as the optimization of factors, models and features, quality and competitiveness. It is a current powerful technique, indispensable in any experience, either in the definition of data to study – what type of data and how much data, or to choose the method and conditions of gathering the samples, always looking for the maximization of feedback information and minimizing costs.
Experimental Design applications are known from experiments in areas as diverse as Medicine, Engineering, Cryptography, Bioinformatics, Social Sciences and Education Sciences.
The technological innovations of today allowed prodigious advances in all areas of research and in particular at the level of Statistics and Experimental Design. Besides the usual computer programs such as STATISTICA, SPSS and SAS, with a relevant role in the programs of classroom teaching, researchers and teachers felt the need to create simple software, free, open to the community and manageable according to the specific needs in each case. R emerges as the current program for more investment in the scientific community of Statistics, making it especially attractive in education programs online. This paper investigates strategies and methods of Experimental Design, as well as R developments, aimed at applications in e-Learning/e-Teaching of these important themes in Masters courses online. Examples of experiences at UAb-Portugal will be presented.

Testing absence of clusters for 1-dimensional observations

Victor Olman, Leonid Morozensky, Ying Hu and Zeev Volkovich

Department of Biochemistry and Molecular Biology, CSBL, University of Georgia, Athens, USA, olman@bmb.uga.edu
ORT Braude Academic College of Engineering, Karmiel, Israel, leon407@013net.net
Department of Biochemistry and Molecular Biology, CSBL, University of Georgia, Athens, USA, xyn@bmb.uga.edu
ORT Braude Academic College of Engineering, Karmiel, Israel, vlvolkov@ort.org.il

The presence of clusters in a sample can be shown through the rejection of the uniformity of the sample's distribution. While many existing techniques deal the asymptotical behavior of statistical tests (such as Pearson χ^2) there is a need for the solution for the comparatively low sample's volume. Such applications include, for example, many problems in bioinformatics, such as microarray analysis, genes concentration along genomes, binding motifs concentration in promoters regions and etc. We present a new statistical test based on sum of squared distances between neighboring 1-dimensional observations. Analytical results for the moments of the statistics are given as well as a computational recursive procedure for P-value calculation. Examples for testing the scattering of onco-suppressor genes along genomes are given.

An application of stochastic dominance in ruin and risk problem

António Pacheco[1] and Elena Almaraz Luengo[2]

[1] CEMAT and Instituto Superior Tecnico, TU, Lisbon, Portugal, apacheco@math.ist.utl.pt
[2] CEMAT and Facultad de Ciencias Matematicas-UCM
UCM, Madrid, Spain, ealmarazluengo@mat.ucm.es

In this work we will consider a semi-Markov risk model

$$X(t) = X(0) + \int c_{J(s)} ds - \sum_{n=1}^{N(t)} Y_n$$

where $\int_0^t c_{J(s)} ds > 0$, $X(0) \geq 0$, being $X(0)$ the initial capital, $J(s)$ a semi-Markov modulated process, c_j the premium density when the process $J(s)$ is in the state j, Yn the size of the n-th claim and $N(t)$ is the number of claims up to time t.

This model is a generalization of the classical risk model due to Cramer and Lundberg, where the premium density is assumed to be constant, $N(t)$ is an homogeneous Poisson process and Y_j are independent and identically distributed and positive random variables.

Markov-modulated risk model has been studied for example, in Lu, Y. and Li, S. (2005) and Albrecher, H. and Boxma, O.J. (2005).

In this work we study the ruin time: $T_u = \inf\{t \geq : X(t) \leq 0\} \mid X(0) = u$, and we set sufficient conditions for the usual stochastic dominance between the ruin times of two processes under consideration, using a similar methodology to Ferreira, F. and Pacheco, A. (2007). Finally we simulate two processes with ruin times verifying the FSD condition.

Keywords: Coupling, Markov chains, semi-Markov processes, stochastic ordering.

References
1. Albrecher, H., and Boxma, O.J., "On the discounted penalty function in a Markov-dependent risk model", *Insurance Mathematics and Economics* 37, 650--672 (2005).
2. Ferreira, F., and Pacheco, A., "Comparision of level-crossing times for Markov and semi-Markov processes", *Statistics and Probability Letters* 77, 151--157 (2007).
3. Lu, Y., and Li,S., "On the probability of ruin in a Markov-modulated risk model", Insurance Mathematics and Economics 37, 522--532 (2005).

Stochastic Clustering and Nucleation

Florentin Paladi

Department of Theoretical Physics, State University of Moldova
A.Mateevici str. 60, Chisinau MD-2009, Moldova
fpaladi@usm.md

In this paper the present developments in the physics of complex systems, in particular the structural relaxation of supercooled liquids and glasses, are discussed by using a stochastic cluster-based model. We are able to depict the impact of the interface between the nucleus considered as a cluster of a certain number of molecules and the liquid phase for the enhancement of the overall nucleation process. In general, these mathematical models describe the interactions of agents in heterogeneous populations and they are developed within the framework of the recent discussions about the gap between agent-based computational models (ABM) and stochastic analytical models. In particular, it is shown that even a relatively simple stochastic model, which appears phenomenological if it is not agent-based, can describe precisely the outcomes from multiple agent-based simulations where there is a lack of probabilistic insight and which should be long enough to equilibrate the states of large systems.

Keywords: Stochastic modeling, Phase transitions; Nucleation; ABM

A New Method for Dynamic Panel Data Models with Random Effects

Savas Papadopoulos

Democritus University of Thrace

A simple-open-form estimator is introduced for the dynamic coefficient and it can be applied to levels. In a dynamic model without additional regressors, a lag of order three is included which handles the random effects. It is shown via simulations that the difference between the two OLS coefficient estimators of order-one and order-three lags estimates consistently the dynamic coefficient. In a model with other regressors, a method is suggested which estimates all the coefficients individually using restricted least squares (RLS). After estimating all the coefficients of the static regressors, the dynamic coefficient can be estimated by restricting the coefficients of the regressors to their estimated values by the suggested method (RLS) or by another method, e.g., transformed MLE (TMLE) or GMM. The RLS method can be applied when the sample size N is relatively small to the number of periods T and when the methods TMLE and GMM cannot be applied. In an application, it is shown that the RLS method provides smaller RMSE's than TMLE and GMM. Simulations compare RLS with TMLE and GMM. In general, RLS performs better than GMM. TMLE gives better results than GMM and RLS in some cases but indicates convergence problems when N and T are small.

Keywords: Restricted regression, transformed maximum likelihood, Arellano-Bond Estimator.

Modeling Biological Sequences and Web Navigation with a Semi Markov Chain

Aleka A. Papadopoulou

Department of Mathematics, Aristotle University of Thessaloniki,
Thessaloniki 54124, Greece, apapado@math.auth.gr

In semi Markov processes we are sometimes concerned with the entrance of the process into a state rather than with the presence of the process in that state. Also because the semi Markov model allows a distinction between the number of time units that have passed and the number of transitions that have occurred we have the opportunity of asking the probability distribution of the number of transitions to a state that occurred through a time interval. In the present the entrance probabilities concerned with the number of transitions and the probability distribution of the number of transitions to a state are studied to provide some answers to questions related to state occupancies. Biological sequences and Web navigation are two cases that initially seem to be different but to a certain extend they do have similarities. Two main aspects of word occurrences in biological sequences are: (a) where do they occur and (b) how many times do they occur. In Web navigation the similar questions are (a) when a node is visited and (b) how many times a node is visited. So, the theoretical results of this study are applied to model these two cases and derive distributions of word location or node occurrence and frequency of occurrences.

Keywords: Semi Markov chains, Entrance probabilities, State occupancies, Biological sequences, Words, Web navigation.

Non-Linear Filtering for Telescoping Markov Chains and Applications to Evolving Images

Andrew Papanicolaou[1] and Boris Rozovsky[2]

[1]Division of Applied Mathematics, Brown University, Providence RI, 02912,
andrew_papanicolaou@brown.edu
[2]Division of Applied Mathematics, Brown University, Providence RI, 02912,
boris_rozovsky@brown.edu

In this paper we discuss problems related to identification/prediction of behavioural characteristics of an intelligent agent (e.g. a human, a robot, an avatar) based on noisy video footage. The attributes of interest include: patterns of motion (e.g. pose and position), intentions, mood swings etc. Some of the attributes, (e.g. intensions) are not directly observable, and need to be inferred from the other attributes. Others, such as pose and location, are partially observable but the observations are corrupted by noise.

Our general approach to the problem is Bayesian. More specifically, the hierarchical dynamics of the agent behavior is modeled by a telescoping/ layered Markov chain (TMC) which iteratively conditions the distribution of circumstantial attributes on the values taken by more basic ones. For instance, a person's pose is a random field taking values in a set of possible poses. The distribution with which the agent takes a particular pose depends on the agent's intentions, which can be modeled by components of the TMC at another level in the hierarchy.

Our approach to identification is based on nonlinear filtering type algorithms and optimal change-point detection for partially observable TMCs. Generally speaking, nonlinear filtering for partially observable TMCs is a particular case of the HMM algorithm for vector-valued Markov chains. One of the main obstacles to efficient performance HMMs is the curse of dimensionality. To some degree, these problems could be mollified by introduction of particle filters, Rao-Blackwellization, and other methods, but the high dimensionality still remains to be a serious problem. Introduction of TMCs is just another step in the quest for reduction of computational complexity of HMMs. It shall compliment other methods (e.g. particle filters and Rao-Blackwellization) when possible.

Applications of TMC-based nonlinear filtering to analysis of video footage, presented in the paper, demonstrates practical potential of the approach.

Stein's Method and the Equilibrium Distribution Coupling

Erol Pekoz

Boston University

If a random variable can be well coupled with its equilibrium distribution, then the random variable has close to either a geometric or exponential distribution. We show how this principle can be used with Stein's method to obtain approximation error bounds for some queues and branching processes.

Managing analytic processes: a multi agent and knowledge management structures approach

Igor Perko, Miro Gradišar and Samo Bobek

University of Maribor, Faculty of Economics and Business, Maribor, Slovenia
igor.perko@uni-mb.si

To obtain optimal results in a repeatedly used system of multiple predictive models, an advanced management of analytic process is required. To manage the analytic processes we propose a management system combining a multi-agent system (MAS) with a reasoning

capable knowledge management structures (KMS). Its tasks are reacting to the changes in the environment maintaining the quality level of the results in time and providing universal result representation. The required intelligent agents (IA) capabilities and the KMS structure supporting the IA reasoning process of are examined. The proposed system is tested in a real-life case involving the probability of default evaluation in a commercial bank customer portfolio.

Key Words: Predictive Analytics; Business Intelligence; Intelligent Agents; Multi-Agent Systems; Knowledge Management Structures; Probability of Default

Direct vs. indirect sequential Monte-Carlo filters

Yohan Petetin and François Desbouvries

CITI Department and CNRS UMR 5157
Telecom SudParis, 91011 Evry, France
Email: yohan.petetin@telecom-sudparis.eu, francois.desbouvries@it-sudparis.eu

Let (x,y) be an HMC : $p(x_{0:n}, y_{0:n}) = p(x_0) \prod_{i=1}^{n} p(x_i \mid x_{i-1}) \prod_{i=0}^{n} p(y_i \mid x_i)$. The Bayesian filtering problem consists in computing (or estimating) $p(x_n \mid y_{0:n})$ for all n. As is well known, $p(x_n \mid y_{0:n})$ can be computed recursively in time as :

$$p(x_n \mid y_{0:n-1}) = \int p(x_n \mid x_{n-1}) p(dx_{n-1} \mid y_{0:n-1}), \quad p(x_n \mid y_{0:n}) = \frac{p(y_n \mid x_n) p(x_n \mid y_{0:n-1})}{\int p(y_n \mid x_n) p(dx_n \mid y_{0:n-1})} \quad (1)$$

In the Gaussian case, exact solutions are provided by Kalman filtering (KF) techniques. In the general case sequential Monte Carlo (SMC) methods aim at computing a discrete approximation of the continuous measure $p(dx_n \mid y_{0:n})$. More precisely, many efforts have been devoted to the MC approximation of (1), which explicitly computes $p(x_n \mid y_{0:n})$ (or in short $p_{n|n}$) as the recursion $p_{n-1|n-1} \rightarrow p_{n|n-1} \rightarrow p_{n|n}$.

In this paper we do not try to further improve MC approximations of equations (1); by contrast we focus on (1) themselves, and indeed explore alternate paths for computing $p_{n|n}$ recursively, even if $p_{n|n}$ is obtained (indirectly) as the byproduct of some alternate recursive algorithm. These ideas are well known in the context of KF, but not in particle filtering (PF), so the aim of this paper is to investigate them in the context of SMC techniques. We first explore the alternate path $p_{n-1|n-1} \rightarrow p_{n-1|n} \rightarrow p_{n|n}$. Both solutions differ only by the intermediate step which is either the predictive distribution $p_{n|n-1}$ or the smoothing one $p_{n-1|n}$. These two densities can in turn be propagated via the two different paths which are obtained when moving one time index and next the other. This observation naturally yields six algorithms (four of which are distinct) for computing $p_{n|n}$ recursively : two direct, two prediction-based and two smoothing-based.

It remains to approximate these paths in practice. Since each of them can be decomposed into a Bayesian step and a Markovian step (or vice versa), we propose a common MC routine for implementing the recursions. Finally we get four generic SMC algorithms (two of which are original). We finally compare via simulations the performance of these algorithms and of the SIR algorithm with optimal importance pdf. Simulations show that the smoothing-based SMC filters outperform the prediction-based ones, and that in each class of algorithms better results are obtained when updating precedes propagation.

EM and ICE in Hidden and Triplet Markov Models

Wojciech Pieczynski

CITI Department, Institut Telecom, Telecom Sudparis, Evry, France
ojciech.pieczynski@it-sudparis.eu

In this paper we consider two general parameter estimation methods in the case of incomplete data which are the "Iterative Conditional Estimation" (ICE [2, 5]), and the "Expectation-Maximization" (EM [3]). EM uses the principle of likelihood maximization, while ICE uses an estimator from complete data and conditional expectation. For Y observed data and X hidden ones, let $p_\theta(x, y)$ be the probability distribution depending on a parameter $\theta \in R^m$. Besides, let $\hat{\theta}(x, y)$ be an estimator of $\theta \in R^m$ defined from complete data (x, y). Both EM and ICE are iterative methods: the EM sequence is defined by $\theta^{q+1}(y) = \arg\max_\theta E[\log[p_\theta(X,Y)|Y = y, \theta^q]$, and the ICE one is defined by $\theta^{q+1}(y) = E[\hat{\theta}(X,Y)|Y = y, \theta^q]$.

Delmas compared the both methods in the case of exponential models and independent observations; in particular, he showed that the EM can be seen as a particular ICE in the sense that the former is independent from the used parameterization while the latter is [2]. The aim of the present paper is to extend these results to the classical hidden Markov models (HMM) and the general triplet Markov models (TMM). We will pay particular attention to the case of hidden semi-Markov chains [1], which are particular TMM [4]. In particular, we discuss the difficulty of applying EM in the general TMM context and we show why ICE is easier to use.

References

1. Barbu, V. S., and Limnios, N., *Semi-Markov chains and hidden semi-Markov models toward applications*, Lectures Notes in Statistics, Springer (2008).
2. Delmas, J.-P., An equivalence of the EM and ICE algorithm for exponential family, *IEEE Trans. on Signal Processing*, **45**, 10, 2613-2615 (1997).
3. McLachlan, G. J. and Krishnan, T., *EM Algorithm and Extensions*, Wiley (1997).
4. Pieczynski, and Desbouvries, F., On triplet Markov chains, *In Proc. of ASMDA 2005*, Brest, France (2005).
5. Pieczynski, W. Sur la convergence de l'estimation conditionnelle itérative, *Comptes Rendus Mathématique*, **346**, 7-8, 457-460 (2008).
6. Yu, S.-Z. and Kobayashi, H., Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden Markov model, *IEEE Trans. on Signal Processing*, **54**, 5, 1947-1951 (2006).

Dynamic routing combined to forecast the behavior of traffic in the city of São Paulo using Neuro Fuzzy Network

Ricardo Pinto Ferreira, Carlos Affonso and Renato José Sassi

Universidade Nove de Julho, São Paulo, Brazil, kasparov@uninove.edu.br
Universidade Nove de Julho, São Paulo, Brazil, carlos.affonso@ajato.com.br
Universidade Nove de Julho, São Paulo, Brazil, sassi@uninove.br

The challenge of getting and keeping customers drives the development of new ways to meet the needs of consumption increasingly tending to micro-segmentation of product and market. The new consumption habits of Brazilians have brought products to market with a life cycle shorter therefore increasing volumes of discarded items and aimlessly after use. The objective of this paper is to support the development of a dynamic routing system supported by the behavior of traffic in the city of São Paulo using Neuro Fuzzy Network trained with notable events occurred in city traffic in Area Maximum Restriction of Movement of city of Sao Paulo. The methodology of the work is to study the city of Sao Paulo, more precisely in Area Maximum Restriction of Movement metropolitan area with a concentration of commerce and

business. A survey on the behavior of daily traffic over São Paulo. Through the three-level hierarchical routing is possible to consider not only the basic factors of scripting but also the external factors that directly influence the flow of traffic and the disruption that can be avoided in large cities across gaps in the path (dynamic routing). Predicting the behavior of traffic represents the strategic level of scripting, the dynamic routing represents the tactical level and routing algorithms to the operational level. Figure 1 illustrates the hierarchy of routing proposed for efficient routing in large cities.

Keywords: The behavior of daily traffic, Dynamic Routing, Neuro Fuzzy Network.

The Reconstruction of Quasioptimal Activation Signals out of Noise and Neuronal Nets Investigation

Andrey V. Polovinkin and Aleksey K. Kryukov

Department of Radiophysics, Nyzhny Novgorod State University, Gagarin ave, N.Novgorod, Russia, polovinkin@rf.unn.ru
Department of Calculational Mathematics and Cybernetics, Nyzhny Novgorod State University, Gagarin ave, N.Novgorod, Russia, alkryukov@gmail.com

We consider the problem of reconstruction the signal leading to activation excitable neural or cardiac cells and provided by minimization signal's energy. The efficiency of this method based on recording time series of white Gaussian noise realizations preceding the system's activation is proofed on examples of FitzHugh-Nagumo, Hodgkin-Huxley and Luo-Rudy models in small-noise case. The criteria of distinction small-noise and big-noise cases are proposed. The theoretical ground of this method was the prehistory reconstruction approach. [1].

The relationship between eigenvalues in the system of linearized Hamiltonian equations describing optimal paths and the eigenvalues of excitable system linearized near the initial steady state is found.

The correspondence between quasioptimal activation signal and the stimulus obtained by reverse correlation technique [2] is under consideration. The informational advisability to use signals, obtained by proposed method in the non-small noise case is discussed.

It is shown, that it is possible to investigate the structure of small net of excitable elements employing their known quasioptimal activation signals.

Keywords: Large fluctuations, Reconstruction, Neural nets.

References

[1] Dykman M.I., McClintock P.V.E., Smelyanskiy V.N., Stein N.D., and   Stocks N.G. "Optimal paths and the pre-history problem for large fluctuations in noise-driven systems", *Phys. Rev. Lett.*, **68,** 2718-21 (1992).

[2] Gerstner W. and Kistler W.M. "Spiking Neuron Models. Single Neurons, Populations, Plasticity". Cambridge University Press, 2002.

The Assessment of the Degree of Soil Pollution by Trace Elements Using Statistical Methods

Nicole Pop[1], Gabriela Oprea[2], Cristina Mihali[2], Angela Michnea[3], Marin Senila[4] and Claudia Butean[2]

[1] Department of Mathematics and Computer Sciences, North University of Baia Mare, Baia Mare, 430072, Romania, nic_pop2002@yahoo.com
[2] Department of Chemistry-Biology, North University of Baia Mare, Baia Mare, Romania
[3] Environmental Protection Agency of Maramures County, Baia Mare, Romania
[4] INCDO-INOE 2000-Research Institute for Analytical Instrumentation, Cluj-Napoca, Romania

Soil pollution by trace elements is a major environmental problem since trace elements are highly toxic, have low biodegradability and tend to be cumulative. In the area of the Baia Mare town located in northern-west Romania, the level of soil pollution by trace elements is due

mainly to mining and metallurgical industry operations, but also to a natural background (volcanic rocks) and to an intensive traffic road. Trace elements are omnipresent pollutants with toxic potential even at low levels of exposure. The cumulative toxic effect is higher than the sum of the effects due to a unique trace element. The simultaneous exposure to more trace elements generates genotoxic effects. Co –exposure to more trace elements (cadmium, lead, arsenic, etc.) can produce interactive genotoxic effects even if the concentrations do not exceed the maximum values allowed by the local legislation. Samples of soil from the residential area of Baia Mare were analyzed and the concentration of the following trace elements was determined: copper, zinc, cadmium, lead, nickel, chromium, mercury, arsenic, antimony and tin. The data was statistically analyzed following the co-variance of the variables. The clustering method and the iso-concentrations maps technique were used. For a global assessment of the extent of soil pollution, soil pollution indices were calculated in every location by reporting the concentration of every considered trace element to the alert threshold value and by summing those reports for all the trace elements measured.

References
1. Cook C.M, Sgardelis S.P., Pantis J.D. and Lanaras T., Concentrations of Pb, Zn and Cu in *Taraxacum* spp. in relation to urban pollution., *Bull. Environ. Contam. Toxicol.,* 53, 204-210 (1994).
2. Sawidis T., Marnasidis A., Zachariasidis G. and Stratis J., A study of air pollution with heavy metals in Thessaloniki city (Greece) using trees as biological indicators, *Arch.Environ. Contam. Toxicol.*, 28,118-124 (1995).
3. Tack F.M. and Verloo M.G., Metal contents in stinging nettle (*Urtica dioica* L) as affected by soil characteristics, *Sci. Total Environ*., 192, 31-39, (1996).
4. Ward N.I. and Savage J.M., Metal dispersion and transportational activities using food crops as bio monitors, *Sci. Total Environ*., 146/147, 309-319, (1994).

Some statistical results on multiple type contact process in competition

Pumo Besnik

Agrocampus Ouest, INHP, Angers, France

A contact process with two species in coexistence is given in works of Durrett [2] and Chan et al. [1] for instance. Following ideas in a previous work with Guyon [3], we present some estimation results for these new models.

References
Chan B., Durrett, R., Lanchier, N. (2009) Coexistence for a multiple type contact process with seasons. *Ann. Appl. Prob. 19, 5, 1921-1943.*
Durrett, R. (2009) Coexistence in stochastic spatial models, *Ann. Appl. Prob. 19, 2, 477-496.*
Guyon, X., Pumo, B. (2007) Space time estimation of a particle system model. *Statistics, 41, 5, 395-407.*

High school schedule creation, optimization problems and solution

Lina Pupeikienė

Vilnius Gediminas Technical University, Saulėtekio al. 11, LT – 10223 Vilnius, Lithuania
Lina.Pupeikiene@gmail.com

The timetable problem, searching the timetable for the class assignment in the schools, belongs to the group of NP hard and NP complete problems. In Lithuanian high schools every pupil can choose a lot of subjects by his wish. The problem is more complicated when the every pupil has possibilities to choose not only subjects, but hour per week of this subject too.

However, as the number of teachers, number of pupils, number of different subjects, number of different subject hours, time slots, and the constraints increases, the required time to find at least one feasible solution grows exponentially. Global optimization algorithms are a quite common approach to solve this problem. In this paper, we describe the advantages of distributed school schedule optimization a software system was developed using Java technology and grid computing techniques. Optimization algorithms used in the software include the Monte-Carlo local optimization algorithm, the Simulated Annealing and Bayes global optimization algorithms.

Key Words: Global optimization, school schedule creation, distributed schedule optimization, Monte-Carlo, Simulated Annealing, Bayes.

Modelling of high speed railroad geometry ageing as a nonlinear autoregressive process.

Ing. L. M. Quiroga and Prof. Dr.-Ing. Dr. h. c. E. Schnieder

Institute for Trafic Safety and Automation Technologies, Technische Universität, Braunschweig, Braunschweig, Germany, quiroga@iva.ing.tu-bs.de e.schnieder@tu-bs.de

Travelling safely and comfortably on high speed railway lines requires excellent conditions of the whole railway infrastructure in general and of the railroad geometry in particular. The maintenance process required to achieve such conditions is largely complex and expensive, demanding an increased amount of both human and technical resources. Figure 1 shows the measurements of the longitudinal leveling (NL) for a 200 m track sector for the last 20 years. The NL parameter is representative of the longitudinal mean deviation of the track in respect of the ideal position. If NL exceeds a certain value, the travelling speed on that sector must be reduced. For this reason the French railway operator SNCF has been measuring periodically the geometrical characteristics of its high speed network since its commissioning, i.e. for more than 20 years now.
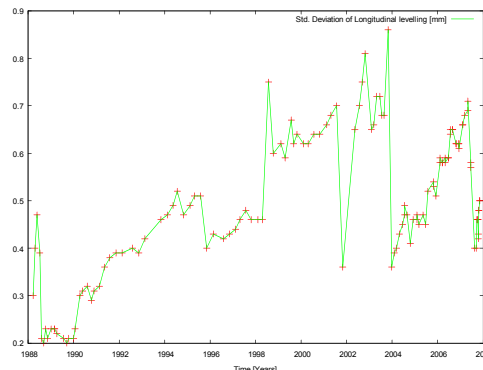
Figure 1: Longitudinal levelling for a railroad sector.

In this framework, choosing the right maintenance strategy is a very important issue. Therefore, a simulation environment in which the different possible maintenance strategies could be tested would be of great advantage. A fundamental requirement for such simulation is a mathematical model describing the behaviour of the railroad geometry deterioration and the effects of the maintenance activities. In this paper a system identification method to obtain such a model is presented. The proposed method uses a "grey-box" model: a model structure and its constraints are specified basing on previous knowledge of the process to be identified, then the set of parameter values which best fits the signal measurements is searched. As previous knowledge indicates that the process is non linear the parameters are searched by means of the Levenberg-Marquardt (LM) algorithm, an iterative technique that finds a minimum of a function expressed as the sum of squares of nonlinear functions. Finally, the method is applied and validated with real data of a French high speed TGV line.

Keywords: Railway Infrastructure, Maintenance, Stochastic modeling, Nonlinear Autoregressive Models.

A scenario tree approach to Optimal Pharmaceutical Research and Development Portfolios using real option valuation method

Majid Rafiee and Farhad Kianfar

Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran,
ma_rafie@mehr.sharif.edu
Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran,
fkianfar@sharif.edu

R&D managers in a pharmaceutical company's pipeline must consider the entire portfolio in the face of market and technological uncertainty and resource constraints. We address the pharmaceutical pipeline management problem by viewing it as a multistage stochastic portfolio optimization. A Multistage stochastic programming approach is introduced for selecting the optimal product portfolio from a set of candidate drugs at different stages in the developmental pipeline and subject to varying levels of market and technical uncertainty over the desired planning horizon. Multistage stochastic programs are effective for solving long-term planning problems under uncertainty. These models use a set of scenarios and corresponding probabilities to model the multivariate random data process (costs or revenues, available budget, chance of success, etc). For most practical problems, the optimization problem that contains all possible scenarios is too large. Due to computational complexity, this program is often approximated by a model involving a (much) smaller number of scenarios. The scenario reduction algorithms determine a subset of the initial scenario set and assign new probabilities to the preserved scenarios by modifying the tree structure and bundling similar scenarios.

This paper presents a stochastic optimization model of pharmaceutical multi-period project selection problem using a real option valuation method. Then it generates related scenario fan and, via existing algorithms, reduces scenarios to generate a new (smaller) scenario tree. Model with reduced scenarios has been solved in this paper. Numerical results indicate improved portfolios and show the effectiveness of the proposed algorithm.

Keywords: project selection, scenario tree, stochastic optimization

Data Mining Application for Eutrophication Control in Surface Waters

ANPALAKI J. RAGAVAN

Departments of Mathematics and Statistics, and Civil and Environmental Engineering, University of Nevada, Reno, NV 89557 (ragavan@unr.edu)

High total phosphorus concentrations(TP) has been found to be the major cause for Eutrophication and the subsequent depletion of dissolved oxygen(DO), enhancement of dissolved organic carbon(DOC), and poor water quality in Truckee River in Nevada. Identifying the exact pattern of relationship among multiple independent variables related to high TP levels is important to implement remediation methods. In this study non-linear mixed (NLMIXED) modeling, time series cross sectional regression, and non-linear least squares (NLLSQ) minimization were used to model the relationship of multiple independent variables to TP as closely as possible. Independent variables included alkalinity, total soluble phosphorus concentrations(STP), stream flow(SF), water pH, water temperature (Temp), DOC, and DO, sampled monthly at the same time of TP (from January 1997 to December 2007) over six monitoring sites (McCarran Bridge(MB), Wordsworth Bridge(WB), Steamboat Creek(SC), Derby Dam(DD), Lockwood(LW), and North Truckee Drain(NTD)) along Truckee River in Nevada. Seasonal variations and man-made intervention in TP were included in the analysis. Fitted NLMIXED model closely predicted observed data explaining 96.7% of total variation (R^2 =0.908). All independent variables influenced TP significantly at 1% significance level. All six sites contributed significantly towards overall TP at 5% significance level (p<0.0001). NLLSQ minimization solution (0.0694 mg/L) to TP was much above the observed overall minimum (0.001 mg/L). Although much lower mean TP were observed at sites MB, DD, and LW, compared to SC and NTD, smaller variations in TP made them significant

contributors towards overall TP. Non-linearity in the relationship of TP to independent variables significantly influenced the prediction of TP hence should be included in all analyses related to prediction of TP in Truckee River. TP in Truckee River is also subject to significant seasonal fluctuations and man-made interventions. Non-linear programming is a suitable and accurate method to identify possible ranges of TP in Truckee River.

FITTING GENERALIZED LINEAR MIXED MODELS FOR NON NORMAL DATA WITH INTEGRAL APPROXIMATION

ANPALAKI J. RAGAVAN

Departments of Mathematics and Statistics, and Civil and Environmental Engineering, University of Nevada, Reno, NV 89557, ragavan@unr.edu

Application of generalized linear mixed models (GLMM) for non normal data faces several constrains before an objective function can be computed such as, difficulty in constructing a valid joint distribution, dependency of mean and variance between non-normal data which places constrains on possible correlation models for valid joint distributions and desired conditional distributions. One approach to solve these problems is to approximate the log likelihood of the GLMM through integral approximation and to submit the approximated objective function to numerical optimization. In this study integral approximation was used to fit a GLMM for non-normal time series water quality data with multiple correlated independent variables collected over a 10 year period. Maximum likelihood estimation was performed using Laplace and adaptive Gauss-Hermite quadrature (QUAD) approximation methods. Parameter estimation through Laplace approximation added restriction with respect to the R-side covariance structure. QUAD approximation required clustered data hence data were processed by subjects before fitting the model with QUAD approximation. Integral approximation yielded a true objective function which could easily be solved through numerical optimization, and potentially unbiased estimates, although including crossed random effects and multiple subject effects in the model were difficult, and R-side covariance structures were not possible.

Neural Networks vs Genetically Optimized Neural Networks in Time Series Prediction

Ion Railean[1,2] Sorin Moga[1] Monica Borda[2] Cristina Stolojescu[1,3]

1 Institut TELECOM; TELECOM Bretagne, UMR CNRS 3192, Lab-STICC; Université européenne de Bretagne, France, [firstname.lastname@telecom-bretagne.eu]
2 Technical University of Cluj-Napoca, Romania, Faculty of Electronics, Telecommunications and Information Technology, [firstname.lastname@com.utcluj.ro]
3 Politehnica University of Timisoara, Romania, Faculty of Electronics and Telecommunications [firstname.lastname@etc.upt.ro]

This paper deals with methods for finding the suitable weights in an Artificial Neural Network (ANN) using Genetic Algorithms (GA).We study the weakness and strengthness of the proposed approach in case of a statistical data forecasting. We describe a different approach when using the input data during optimization phase. Besides GA, we applied stationary wavelet transform (SWT) as a signal preprocessing, and time-delay neural networks (TDNN) approach for the system's inputs. Our results show that this optimization is suitable only for certain purposes in case of a statistical data prediction.
Keywords: Genetic Algorithms, Artificial Neural Networks, forecasting.

Autocopulas: investigating the interdependence structure of stationary time series

Pál Rakonczai; László Márkus; Andras Zempléni

Dept. Probability Theory and Statistics, Eötvös Loránd University, Budapest, Hungary
rakonczai.p@gmail.com

Here we present a novel approach to the description of the lagged interdependence structure of stationary time series. The idea is to extend the use of copulas to the lagged (one-dimensional) series, to the analogy of the autocorrelation function. The use of such autocopulas can reveal the specifics of the lagged interdependence in a much finer way. However, the lagged interdependence is resulted from the dynamics, governing the series, therefore the known and popular copula models have little to do with that type of interdependence.

True though, it seems rather cumbersome to calculate the exact form of the autocopula even for the simplest nonlinear time series models, so we confine ourselves here to an empirical and simulation based approach. The advantage of using autocopulas lays in the fact that they represent nonlinear dependencies as well, and make it possible e.g. to study the interdependence of high (or low) values of the series separately. The presented methods are capable to check whether autocopulas of an observed process can be distinguished significantly from the autocopulas a of given time series model. The proposed approach is based on the Kendall's transform which reduces the multivariate problem to one dimension.

After illustrating the use of our approach in detecting conditional heteroscedasticity in the AR-ARCH vs. AR case, we apply the proposed methods to investigate the lagged interdependence of river flow time series with particular focus on model choice based on the synchronized appearance of high values.

Key Words: copula; stationary time series; autocopula; goodness-of-fit test; Kendall's transform; river flow series

Copula Fitting to Time-Dependent Data, with Applications to Wind Speed Modelling

Pál Rakonczai[1], László Varga[2], and András Zempléni[3]

[1] *Eötvös Loránd* University, Department of Probability Theory and Statistics, Budapest, Hungary

[1] paulo@math.elte.hu, [2] zempleni@math.elte.hu, [3] vargal4@upcmail.hu

Copulae became a popular tools in multivariate modelling, with several fitting methods readily available. General copula models are based on independent samples. In an earlier paper Rakonczai and Zempl´eni [2] focused on the goodness of fit for copulae.

To assume complete independence for a given sample obtained from a time series data is usually too optimistic. Now, as in real applications time dependence is a common feature, we turn to the investigation of the effect of this phenomenon to the proposed test-statistics, especially to the Kendall's process approach of Genest et al (2006, [1]). The block-bootstrap methodology is used for defining the effective sample size for time dependent observations. The critical values are then computed by simulation from independent samples with the adjusted size, determined by these bootstrap methods.

The methods are illustrated by simulated examples and by 2-dimensional models for 50-years long observations of wind data for German cities.

Keywords: copula, goodness of fit, Kendall's process, block bootstrap, effective sample size.

References
1. Genest, C. Quessy, J.-F. and Rémillard, B. (2006) Goodnes-of-fit Procedures for Copula Models Based on the Integral Probability Transformation, Scandinavian J. of Statistics, 33, 337-366.
2. Rakonczai, P. and Zempleni, A. (2007). Copulas and Goodness of fit tests, Recent Advances in Stochastic Modeling and Data Analysis, Chania, Crete, 198-206.

Stochastic movement models generating circular distributions.

William J. Reed

Dept. of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada,
reed@math.uvic.ca

A number of stochastic movement models, leading to circular (angular) distributions are discussed. They can be divided into two categories: (i) models in which the state of the system is an angle (direction or orientation); and (ii) models in which the state of the system is a point in the plane. Various stopping rules are considered. Resulting angular distributions include the wrapped versions of the normal, stable, Laplace and normal-Laplace (NL) distributions; the projected normal and von Mises distribution, and a new skewed form of the von Mises distribution. The fit of some of the models to angular data will also be discussed.
Keywords: Angular distribution; directional data; movement models; Brownian motion; Ornstein-Uhlenbeck process.

Grouping Ordinal Variables by Using Fuzzy Cluster Analysis

Hana Řezanková, Dušan Húsek, Michaela Ryšánková

Dept. of Statistics and Probability, University of Economics, Prague, Czech Republic
hana.rezankova@vse.cz, michaela.rysankova@vse.cz
Institute of Computer Science, Academy of Sciences of the Czech Republic,
dusan.husek@cs.cas.cz

In the data mining tasks, objects are usually characterized by features sets encoded as vectors. Because of its usual high-dimensionality, it is desirable to reduce its dimension before data analysis.

In the paper we study searching groups of similar variables which are of ordinal type. As example we analyze the data from the questionnaire survey on "Active lifestyle of university students". For the investigation purposes 15 variables expressing a satisfaction concerning different points of view of the students' life were included. Respondents expressed their satisfaction using scale from 1 (no satisfaction) to 7 (very satisfied). Answers from 1,453 respondents were collected.

Here we suggest application of selected methods for dimensionality reduction followed by the obtained results interpretation using fuzzy cluster analysis. We apply non-metric multidimensional scaling on the basis of Kendall's tau-b for creation the similarity matrix and categorical principal component analysis. For the analyses SPSS, STATISTICA and S-PLUS systems are used. We determine the optimal number of clusters by using average silhouette widths. The soft version of CSPA (cluster-based similarity partitioning algorithm) is applied for ensembles of fuzzy clustering results obtained on the basis of different techniques.

Suggested approach merges results obtained by several dimensionality reduction methods. The final result contains membership degrees of the ordinal variables to the determined number of groups. The assignment of variables to these groups can be graphically presented by a silhouette plot.
Keywords: Dimensionality reduction, Ordinal variables, Multidimensional scaling, Categorical principal component analysis, Fuzzy cluster analysis

Arbitrage opportunities between NYSE and XETRA?: A comparison of simulation and high frequency data

Jörg Rieger, Kirsten Rüchardt and Bodo Vogt

Faculty of Economics and Management, Chair of Empirical Economics
Otto-von-Guericke-University Magdeburg, Magdeburg, Germany,
Kirsten.Ruechardt@ovgu.de

This paper investigates the no-arbitrage condition of financial markets by comparing two stock markets: the New York Stock Exchange (NYSE) and the German Exchange Electronic Trading System (XETRA). We analyze German stocks that are traded simultaneously at both exchanges using high frequency data for XETRA, the NYSE, and the foreign exchange rates. Converting Euro-prices into Dollar-prices and vice versa reveals possibilities to discuss the efficiency of these two stock markets and arbitrage opportunities. One measure of efficiency is stock price clustering and we obtain the result that XETRA is more efficient if the exchange rate is taken into account. The observed difference in the clustering effect would not be observable, if the no-arbitrage condition held. We propose a trading strategy that exploits these differences. Furthermore, we compare our empirical findings with the results we obtain by simulating financial markets using a Random Walk as a model for the price movement.
Keywords: financial markets; simulation; no-arbitrage condition; stochastic processes

Generalizability Analysis: An Example using Unbalanced Data

Teresa Rivas and Rosa Bersabé

Psychology Faculty. Málaga University. Málaga. Spain, moya@uma.es, bersabe@uma.es

Generalizability Theory (GT) was developed by Cronbach et al. (1972). Only a small number of applications - in educational testing, marketing, etc. - have appeared (e.g., Brennan et al., 1995; Finn, 2004; Bruckner, et al., 2006). TG Analysis was restricted to balanced and non-missing data. Brennan (2001a) has provided extensions of TG models which are better adapted to real situation measurements in psychological and educational practice (unbalanced, missing data, etc). Software is also now available (Brennan, 2001b).
This paper shows an application of a GT model for unbalanced data, the design being $p^\bullet \times i^\circ$, p (subjects) crossed with the fixed facets (scales) and i (items) nested in each fixed facet.
The data are answers given by 778 females to the questionnaire, Eating Attitudes Test (EAT-26; Garner, Olmsted, Bohr & Garfinkel, 1982). Generalizability Coefficients obtained in Generalizability and Decision Studies are described and interpreted. From these results a composite score across scales is also given.
Keywords: Generalizability, Unbalanced data, Composite Score

References
Brennan, R.L. (2001a). *Generalizability Theory*. Springer: New York.
Brennan, R.L. (2001b). *Manual for mGENOVA*. Iowa Testing Programs Occasional Papers No. 50.
Brennan, R.L., Gao, X. & Colton, D.A. (1995). Generalizability analyses of work keys listening and reading. *Educational and Psychological Measurement 55*, 157–176.
Bruckner, C. T.,Yoder, P.J. & McWilliam, R.A. (2006). Generalizability and Decision Studies: An Example Using Conversational Language Samples. *Journal of Early Intervention 28*, 139-153
Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972). *The Dependability of Behavioural Measurements: Theory of Generalizability for Scores and Profiles*. Wiley: New York.
Finn, A. (2004). A reassessment of the dimensionality of retail performance: a multivariate generalizability theory perspective. *Journal of Retailing and Consumer Services, 11,* 235-245
Garner, D.M., Olmsted, M.P., Bohr, Y., & Garfinkel, P.E. (1982). The Eating Attitudes Test: psychometric features and clinical correlates. *Psychological Medicine, 12*, 871-878.

Procedure to Calculate Efficiency Assessment of Three-Category Classification

Teresa Rivas and Félix Caballero

Psychology Faculty. Málaga University. Málaga. Spain, moya@uma.es, ffcaballero@uma.es

An efficacy index is defined to show the utility of a classifier when two cut-off points and three classes are considered.

Given two cut-off points, a scoring classifier $\vec{x}$ is determined by the point (x_1, x_2, x_3, x_4, x_5, x_6) which represents the possible six types of errors. A graphical representation of $\vec{x}$ in 3-D is given as (Fa_R, Fb_R, Fc_R), Fi_R being the proportion of cases whose real state is i ($i = a, b, c$) although they have been classified in classes other than i.

Given any classifier $\vec{x}$ in the unity cube, with $\vec{x}$ under-lying the plane $Fa\_R + Fb\_R + Fc\_R = 2$, the index based on the volume of tetrahedron (IVT) is defined as

$$IVT = 1 - \frac{Fa\_R + Fb\_R + Fc\_R}{2} \quad \text{being} \quad 0 \le Fa\_R + Fb\_R + Fc\_R \le 2$$

The procedure to obtain IVT is shown and its properties are described. An example is also given.

Keywords: Measurement, Multiple Cut-offs, Three-Class Classification, Efficacy Assessment

References

Caballero, F. F. & Rivas, T. (2009, September). Indices de Eficacia en un clasificación tricotómica (I). Paper presented at XII Conferencia Española de Biometría (CEB2009). Cádiz University, pp.:75-76

Ferri, C., Hernández-Orallo, J., Salido, M.A. (2003) Volume under the ROC Surface for Multi-class Problems. In N. Lavrac, D. Gamberger, L. Todorovski, & H. Blockeel (eds.): *Proc. of the 14th European Conference on achine Learning* (pp. 108–120). Cavtat-Dubrovnik, Croatia: Springer

Ferri, C. Hernández-Orallo, J. & Salido, M.A. (2003) Volume Under the ROC Surface for Multi-class Problems. Exact Computation and Evaluation of Approximation. Technical Report. 1-40. Retrieved May 20, 2009, from http://users.dsic.upv.es/grupos/gps/papers/VUS-report.pdf

Rivas, T. & Caballero, F. F. (2009, September). Indices de Eficacia en un clasificación tricotómica (II). Paper presented at XII Conferencia Española de Biometría (CEB2009). Cádiz University, Pp.:77-78

Stein Couplings and What We Can Do With Them

Adrian Roellin

University of Singapore

Stein's method, introduced in the early 70's by Charles Stein, is an approach to obtain distributional approximations, concentration of measures and related results. One of the key steps to successfully implement the method is to establish a so called 'Stein –identity' for the random variable under investigation. Once such an identity is found several interesting results can be deduced from it. However, it remains often unclear how one can construct can construct such an identity for a specific problem. Introducing the new concept of 'Stein couplings' we provide a more systematic approach to obtain Stein identities. We give very general theorems for normal approximation in the Wasserstein and Kolmogorov metrics and, once a Stein coupling is found, the final bound can be obtained through bounding some moments involving the coupled random variables. We provide new examples from random graph theory, urn models and neighborhood statistics of random points on the d-dimensional cube.

Hidden Markov Models for Financial Optimisation

Diana Roman, Gautam Mitra and Nicola Spagnolo

Brunel University, City Uxbridge, UK, diana.roman@brunel.ac.uk

Many financial decision problems require scenarios for multi-variate financial time series that capture their sequentially changing behaviour, including their extreme movements. We consider modelling financial time series by Hidden Markov models, which are regime switching type models. Estimating the parameters of a HMM is a difficult task and the multi-variate case can pose serious implementation issues. After the parameter estimation, the calibrated model can be used as a scenario generator to describe the future realisations of asset prices. The scenario generator is tested in a single-period mean-CVaR optimisation problem for portfolio selection.
Key Words: scenario generation, asset pricing, Hidden Markov models, tail behaviour, extreme events.

The Multiple Player Ante One Game

Sheldon M. Ross

University of Southern California

Consider a group of players playing a sequence of games. There are k players having arbitrary initial fortunes. Each round consists of each remaining player putting 1 in a pot which is then won (with equal probability) by one of them. Players whose fortune drops to 0 are eliminated. When k=3, we use martingale stopping theory to derive both the expected number of games played until one of the players has all the money, and the expected number of games played by a specified player. When k is larger than 3 we derive bounds.

Evaluation of the Mean Squared Error of the Estimate of the Asymmetry Parameter in Stable Law

Viacheslav V. Saenko and Olga Yanushkevichiene

Ulyanovsk State Universite, Ulyanovsk, Russia, saenkovv@gmail.com
Institute of Mathematics and Informatics and Vilnius Pedagogical University, Vilnius, Lithuania
olgjan@zebra.lt

Let $X_j, j = 1,...,3n$ be independent identical distributed stable random variables with characteristic function

$$f(t; \alpha, \beta, \gamma, \lambda) = \begin{cases} \exp\left(\lambda\left(it\gamma - |t|^\alpha + i|t|^\alpha \beta \tan(\pi\alpha/2)\mathrm{sign}\,t\right)\right), & \alpha \neq 1 \\ \exp\left(\lambda\left(it\gamma - |t|^\alpha + i\beta(2/\pi)\log|t|\right)\right), & \alpha = 1 \end{cases}$$

where $0 < \alpha \leq 2, -1 \leq \beta \leq 1, |\gamma| < \infty, \lambda > \infty$.

The estimation algorithm of the parameters α, β, λ was described in [1, 2]. It based in transition from representation of the characteristic function $f(t; \alpha, \beta, \gamma, \lambda)$ to representation

$$f(t; \nu, \theta, \tau) = \exp\left(-\exp\left(\nu^{-1/2}\left(\log|t| + \tau - i(\pi/2)\theta\,\mathrm{sign}\,t\right) + C(\nu^{-1/2} - 1)\right)\right),$$

by transformation $Y_i = X_{3i-2} - 1/2(X_{3i-1} + X_{3i})$, $i = 1,...,n$. Here $C = 0.577...$ is the Euler constant and the parameters change in the limits $\nu \geq 1/4, |\theta| \leq \min(1, 2\sqrt{\nu} - 1), |\tau| < \infty$. The estimators of the parameters ν, θ, τ has the form

$$\overline{\nu} = \max(\hat{\nu}, 1/4), \quad \overline{\theta} = \min(1/3, |\hat{\theta}|)\operatorname{sign}\hat{\theta}, \quad \hat{\tau} = A_V \,,$$

where $U = \operatorname{sign} Y_i$, $V = \log|Y_i|$, $\hat{\nu} = \left(6/\pi^2\right)B_V - (3/2)B_U + 1$, $\hat{\theta} = A_U$ and

$$A_W = \frac{1}{n}\sum_{i=1}^{n}W_i, \quad B_W = \frac{1}{n-1}\sum_{i=1}^{n}(W_i - A_W)^2$$

For estimation of mean squared error of the asymmetry parameter of stable law the following theorem is valid

Theorem. *Under the assumption that* $\overline{\nu} \geq \nu_0 > 1/4$ *the estimate* $\overline{\beta} = \min(|\hat{\beta}|, 1)\operatorname{sign}\hat{\beta}$ *of the parameter* β *, where*

$$\hat{\beta} = \begin{cases} \dfrac{1 + 2^{1-1/\sqrt{\nu}}}{1 - 2^{1-1/\sqrt{\nu}}} \cdot \dfrac{\tan\left(\pi\overline{\theta}/2\sqrt{\nu}\right)}{\tan\left(\pi/2\sqrt{\nu}\right)}, & \nu \neq 1 \\[3mm] -\dfrac{\pi}{\log 2}\tan\left(\pi\overline{\theta}/2\right), & \nu = 1 \end{cases}$$

is asymptotically unbiased of order $1/\sqrt{n}$ *, and the following inequalities are valid*

$$E(\overline{\beta} - \beta)^2 \leq \left(K_1\sqrt{E(\hat{\nu} - \nu)^2} + K_2\sqrt{E(\hat{\theta} - \theta)^2}\right)^2, \quad \nu \neq 1,$$

$$E(\overline{\beta} - \beta)^2 \leq \frac{\pi^4}{3\log^2 2}E(\hat{\theta} - \theta)^2 + \frac{2\pi^5}{9\log^2 2}E(\hat{\theta} - \theta)^3 + \frac{\pi^6}{27\log^2 2}E(\hat{\theta} - \theta)^4, \quad \nu = 1.$$

The expressions for $K_1, K_2, E(\hat{\nu} - \nu)^2, E(\hat{\theta} - \theta)^2, E(\hat{\theta} - \theta)^3, E(\hat{\theta} - \theta)^4$ will defined in the report. This result considerable improves result of the work [3].

References
1. Zolotarev V. M., Statistical Estimation of Parameters of Stable Laws. *Banch center publication*, **6**, 359-376 (1980).
2. Zolotarev V. M., One-Dimensinal Stable Distributions. Amer. Mat. Soc., Providence,RI, 1986.
3. Yanushkevichiene O., Investigation of some parameters of stable laws. *Lietuvos mathematikos rinkinys,***21**, 195-209 (1981).

A new Local EM Estimation Method for Latent Factorial Generalized Linear Models

Mohamed Saidane[1], Christian Lavergne[1, 2], and Xavier Bry[1]

[1] Universite Montpellier II, I3M UMR-CNRS 5149, Place Eugene Bataillon CC 051 - 34095, Montpellier, France, saidane@math.univ-montp2.fr, bry@math.univ-montp2.fr
[2] Universite Paul-Valery, Montpellier III, route de Mende - 34095, Montpellier, France, Christian.Lavergne@math.univ-montp2.fr

Factor models have been fully developed and dealt with in the case where observations are assumed to be normally distributed. Here, we consider the less restrictive framework in which the distribution of the observations is assumed to belong to the exponential family. Thus, we introduce a new class of factor models allowing to analyze and predict discrete data (binomial, Poisson...), but also non-normal continuous data (gamma, for instance). These Generalized Linear Factor Models (GLFM) are built up combining standard Factor Models with Generalized Linear Models (GLM). A new parameter estimation method is presented for the GLFM. It is based on Fisher's Score algorithm for non-standard GLM, combined with an Expectation-Maximization (EM) type iterative algorithm for latent factors. Extensive Monte **Carlo simulations show promising results.**
Keywords: Factor Models, Generalized Linear Models, EM Algorithm, Scores Algorithm, Simulations.

Adaptive Monte-Carlo Markov Chains

Leonidas Sakalauskas

Operation Research Department, Institute of Mathematics & Informatics, Akademijos st 4, 08663 Vilnius, Lithuania, sakal@ktl.mii.lt

Monte- Monte-Carlo Markov Chain (MCMC) approach is widely applied to statistical estimation and decision making. However, implementation of MCMC relates with significant computational expenses, which are necessary to realize the method, uncertainty in choice of chain parameters and nonexistence of statistically based termination rules. This paper pursue to develop the adaptive MCMC algorithms that enables us to find statistical estimates with admissible accuracy. Our approach distinguishes by adaptation of MCMC samples size so, that to decrease the total amount of computations and to ensure the convergence to solution of MCMC equations, as well as a treatment of randomization errors in statistical manner. The properties of algorithm developed are demonstrated on statistical estimation of parameterts of multivariate skew t-distribution. Results of computer simulation and solving test examples are disscussed, too.
Keywords: Stochastic modeling, Statistical Simulation, Monte-Carlo Markov Chain, t-skew distribution, Hotelling statistics

Estimation of Skew *t* - Distribution by Monte-Carlo Markov Chain Approach

Leonidas Sakalauskas and Ingrida Vaiciulyte

Institute of Mathematics & Informatics, Vilnius, Lithuania, sakal@ktl.mii.lt, ingrida_vaiciulyte@yahoo.com

The algorithm for application of Monte - Carlo Markov Chain (MCMC) for estimation of parameters of skew *t* - distribution is developed. In general, the skew *t* - distribution is represented by multivariate skew-normal distribution with covariance matrix depending on parameter distributed according to inverse-gamma distribution (Azzalini (1985), Kim and Mallick (2003), Azzalini and Genton (2008)). Using this representation, the density of skew *t* - distribution is expressed through multivariate integral which is convenient to estimate numerically by Monte - Carlo simulation. Thus, the likelihood function, depending on some data sample, might be estimated by Monte - Carlo method, too. Next, the MCMC procedure is developed for recurrent estimation of skew *t* - distribution, where the Monte - Carlo sample size is regulated so that to ensure the convergence and to decrease the total amount of Monte - Carlo trials. The Monte - Carlo estimates of confidence intervals of estimates are introduced using the Gaussian approximation according to the CLT. The termination rule is implemented testing statistical hypotheses about insignificant change of estimates in two steps of the procedure (Sakalauskas (2002)).
Keywords: Statistical simulation, Monte – Carlo method, Maximal likelihood, Gaussian aproximation, EM – algorithm.

Data exploration and analysis for obtaining mortality profiles from the animal bones remains found in archaeological sites

A. Scaringella and G. Siracusano

Facoltà di Sociologia, Università di Roma "Sapienza", Roma, Italy
angela.scaringella@uniroma1.it

We discuss the problem of obtaining mortality profiles from some animal bones remains found in archaeological sites. The method takes account of different types of bones that can be related to different sets of possible ages of the animals. As a first preliminary step we obtain the likelihood function and derive the equations for the point of maximum likelihood.
Key Words: Mortality profile, Likelihood analysis, Archaeozoology

Reconfiguration State Heath Services Logistics: Patients Flow Optimization

Cassius Tadeu Scarpin, Maria Teresinha Arns Steiner and Pedro José Steiner Neto

Federal University of Parana, Curitiba, Paraná Brasil, tere@ufpr.br

This paper presents an optimization to the flow patients for state health services offered to Parana State citizens and also analyzes a new optimized configuration for it. In Brazil, medical care is controlled by state officers, and is provided throughout a centralized planning managed by state governments in every one of Brazil's 28 states. Since geographical area is a key factor in most states, and range of services are not the same for every unit, patients demanding my have to be transported to another facility, located in the same state. A new proposition is made for state division into smaller regions, with a service level assigned for each municipality. Cities with lower levels offer just basic procedures, and patient in need of a more specialized service are sent to higher level cities. The assignment for the groups of cities must be based on size of population, health facilities, road availability and patient transportation routes should be planned for each procedure in every facility. For the regional division, a branch-and-price algorithm was used, using the column generation algorithm for each node in the branch-and-bound tree. The results obtained were considered very consistent.
Key Words: Logistics; p-medians problem; branch-and-price algorithm; Flow of Patients; Hierarchical Configurations.

The Application of the ROC curve to Classification and Regression Trees

M. Sedlacik

Faculty of Economics and Management, University of Defence, Department of Econometrics, 662 10 Brno, Kounicova 65, Czech Republic, marek.sedlacik@unob.cz

The classification and regression trees (CART) methodology plays an important role in many branches, especially when it is necessary to differentiate between two or more populations. The splitting rule in the CART is considered to be the decisive factor of the algorithm. The conventional method of the splitting has been proposed by Breiman et al. (1984).
The contribution is focused on the implementation of the receiver operating characteristic curves (ROC) to the classification and regression tree analysis as the alternative splitting of nonterminal nodes. First, a brief introduction of the CART methodology and ROC curves theory (including basic properties and different techniques of estimation) is mentioned. Finally, two possibilities how to incorporate ROC curves into the CART construction are suggested.
Keywords: Classification and regression trees, CART, Splitting rule, Receiver operating characteristic curve, ROC

Numerical analysis of random signals with singularities

Oleg Seleznjev and Konrad Abramowicz

Department of Math. and Math. Statistics, Umeå University, SE-901 87 Umeå, Sweden
oleg.seleznjev@math.umu.se, konrad.abramowicz@math.umu.se

The close relationship between the smoothness properties of a function and the best rate of its linear approximation is one of the basic ideas of conventional (deterministic) approximation theory. We study similar properties for random signals (processes). In particular, we consider piecewise polynomial approximation of a continuous (in quadratic mean, q.m.) random process $X(t)$, $t \in [0,1]$, based on n observations. The performance of the approximation is

measured by mean errors (integrated or maximal q.m. errors). Let X satisfy a Hölder condition with exponent $0 < a < 1$, say, $H^a[0,1]$, and have a continuous q.m. m-derivative, $m \square 1$, for all points $t > 0$. It is known that the approximation rate n^a is optimal for linear approximation methods in a certain sense for the Hölder class $H^a[0,1]$ (Buslaev and Seleznjev, 1999, Seleznjev, 2000). But for such smooth process with singularity at one point (or a finite number of points) and a certain locally stationarity property, we investigate the sequence of quasi-regular designs (observation locations) and find the sequence of sampling designs with approximation rate n^m and asymptotically optimal properties as $n \square \square$, e.g., for integrated q.m. norm.. These results can be used in various problems in numerical analysis of random functions, for archiving telecommunication, multimedia, environmental data in databases, and in simulation studies.

Keywords: Random signal, Process, Approximation

References

1. Buslaev, A.P. and Seleznjev,O., On certain extremal problems in theory of approximation of random processes. *East Jour. Approx.* **5**, 467-481 (1999).
2. Seleznjev, O., Spline approximation of random processes and design problems. *Jour. Stat. Plan. Infer*. **84**, 249-262(2000).

Construction of algorithms for calculation of a wavelet spectral density estimate

Natallia Semianchuk and Mikolaj Troush

Grodno State University, Grodno, Belarus, senata155@gmail.com
Belorusian State University, Minsk, Belarus, troushNN@bsu.by

Let $X(t)$, $t \in \mathbf{Z}$, be a wide-sense stationary stochastic process with $EX(t) = 0$, $t \in \mathbf{Z}$, belonging to a set of random processes $\chi(\lambda, f, \alpha, L, C_2)$. The set $\chi(\lambda, f, \alpha, L, C_2)$ is defined as the set of wide-sense stationary processes X(*t*), , $t \in \mathbf{Z}$, whose spectral density is $f(\lambda)$, $\lambda \in \Pi = [-\pi, \pi]$, having a fourth-order semi-invariant spectral density $f_4(\lambda_1, \lambda_2, \lambda_3)$, $\lambda_j \in \Pi$, $j = \overline{1,3}$, and such that for fixed $\lambda \in \Pi$ the spectral density f satisfies $f \in Lip_\alpha(L)$ and the fourth-order semi-invariant spectral density is bounded by a constant $C_2 > 0$.

For a process $X \in \chi(\lambda, f, \alpha, L, C_2)$, the rate of convergence of the mean-square deviation of a linear wavelet estimate of the spectral density is studied. The coefficients of the asymptotically dominant term, which depend on the smoothness of the spectral density, are calculated for some scaling functions and data tapering windows.

Our theoretical results are also used for developing computational algorithms for wavelet estimates of the spectral density. These algorithms enable us:

1) to select a data tapering windows,
2) to choose a scaling function,
3) to compute the level of decomposition

in order to construct the estimate minimizing the mean square deviation, depending on the sample length and the smoothness of the spectral density.

Keywords: stochastic process, wavelet spectral density estimate

APPLICATION OF STOCHASTIC MODEL IN GAMMA TYPE FUNCTION FOR PREDICTION OF MILK PRODUCTION IN SAHIWAL CATTLE

R K Sharma

Department of Veterinary & Animal Husbandry Extension, Faculty of Veterinary & Animal Sciences, G B Pant University of Agriculture & Technology, Pantnagar, Uttarakhand-263145 INDIA, rabendra1@rediffmail.com

The milk production is a very important trait in farm animals (cattle & buffaloes) involving various complex physiological and biochemical processes governed by hereditary and environmental factors. Milk production can be expressed in the form of lactation curves which

describes the rate of change of milk secretion with the advance of lactation. The study of lactation curves enables the planners to predict milk production on the basis of part records. It helps animal breeders in sire evaluation, early culling of poor performers and planning of feeding and management of dairy enterprise. Various mathematical models have been used to describe the lactation curves varying from simple exponential function, parabolic exponential function to gamma type function. Keeping the above facts in view, the present investigation was undertaken to predict milk production utilizing the production records of 150 Sahiwal cows and using stochastic model in gamma type function. Sahiwal is the highest milk producing breed of cattle in India. Based on R^2 values the stochastic model in gamma type function (0.990) was found to be the best fit followed by Gamma type (0.978) parabolic (0.935), exponential (0.923), and inverse polynomial function (0.774). t may, therefore, be concluded that the stochastic model in gamma type function may be employed to predict milk yield of farm animals with reasonable accuracy.

Keywords: Cattle, milk production, Sahiwal, stochastic model

Matrix analysis of datebase

Svetlana Sharonova, Anna Sharonova

St.Tikhon's Orthodox University, Moscow, Russia, s_sharonova@mail.ru

According to the Matrix Theory and Graph Theory using patterns it is possible not only to carry out information search, but to carry out the analysis on the basis of texts making a database of library fund. The matrix grid allows to create difficult analytical constructions which parameters it would be possible to translate in other planes of matrix research, thereby modelling and in some measure and predicting development of those or other tendencies of social practice.

The habitual linear content-analysis can be expanded to the nonlinear pattern-analysis when the certain semantic sample from set of texts is created.

At research of social problems the scientist usually addresses to the texts written by sociologists, philosophers and representatives of other humanities, which anyhow to a research related topic. However it is difficultly to assume in what texts of a scientific direction it is possible to come across ideas which are capable to turn developed representation, to change a habitual trajectory of thinking. Speed of social changes is so prompt what to wait for that case when someone will come across these ideas simply not permissible luxury. On the other hand development of computer programs goes with the same stunning speed and artificial intellect possibilities all become more perfect that the sin not to take advantage of them. Matrix researches are already used on-line. On the Internet you can meet on-line web surveys which advertise Matrix Research as providing technical innovation. These programs are calculated on revealing of a client rating.

We suggest to construct the program aimed at research.

Key Words: matrix, content-analysis, pattern

Relation between obesity and depression using structural equation modelling

Sheikhi A.[1], Janghorbani M.[2], Soleimani B.[2]

[1] Center of Biostatistics, University of Limerick, Ireland (ali.sheikhi@ul.ie)
[2] Department of Biostatistics and Epidemiology, Isfahan University of Medical Sciences, Isfahan, Iran (janghorbani@yahoo.com; soleimani_b@yahoo.com)

In the medical sciences, apart from the observation of parametric differences between two groups of subjects, the determination of a network of causes is also desired. In such cases one of the best approaches is Structural Equation Modelling (SEM), especially when a number of variables are not directly measurable. Structural equation modelling can test a series of independent relations simultaneously. This feature is particularly useful when a dependent variable in an equation is converted into an independent variable in another equation. The structural equations are a set of m equations, each having a different

dependent variable and *n* independent variables. The dependent variable in one equation may be an independent variable in another.

By the use of indexes presented by Karl Joreskog, the LISREL pattern can be represented in the following figure:

$$\underset{m\times 1}{\eta} = \underset{m\times m}{\beta}\,\underset{m\times 1}{\eta} + \underset{m\times n}{\Gamma}\,\underset{n\times 1}{\xi} + \underset{m\times 1}{\psi} \qquad \text{as the structural equation model,}$$

and,

$$Y = \Lambda_y\eta + \varepsilon,\ X = \Lambda_x\xi + \delta \qquad \text{as the measurement models for Y and X respectively.}$$

In the present study, the SEM method has been employed to study the association between obesity and depression and a network of variables expected to be linked to these two conditions, on a sample of 1093 inpatients in Khorshid center for behavioral disorders in Isfahan, Iran.

Keywords: SEM, LISREL, Obesity, Depression

New Models for Reselling of Options

Dmitrii Silvestrov and Robin Lundgren

Stockholm University, Stockholm, Sweden, silvestrov@math.su.se
School of Education, Culture, and Communication, Mälardalen University, Västerås, Sweden
robin.lundgren@mdh.se

A problem of optimal reselling of European options is discussed. A bivariate exponential diffusion process with correlated components represented by a geometric Brownian motion and a mean-reverse Ornstein-Uhlenbeck process are used to model the price process and the stochastic implied volatility in the reselling model. The reselling problem is imbedded into the problem of finding the optimal expected reward for American type option based on this bivariate process and a non-standard pay-off function given by the Black-Scholes formula. New skeleton approximations and convergence results for optimal reward functionals of American type options for perturbed multivariate Markov processes are presented. A bivariate binomial-trinomial tree approximation model is constructed and convergence of the optimal expected reward for this tree model to the optimal expected reward for the corresponding American type option is proved with the use of general skeleton and convergence results mentioned above. Results of numerical experiments are presented. Related results and alternative approaches to the problem are also discussed.

Key words: Option, Reselling, Multivariate price Process

Markov and Semi-Markov Stochastically Ordered Models for Credit Rating Dynamics

Dmitrii Silvestrov, Evelina Silvestrova and Raimondo Manca

Stockholm University, Stockholm, Sweden, silvestrov@math.su.se
Mälardalen University, Västerås, Sweden, evelina.silvestrova@mdh.se
Università "La Sapienza", Rome, Italy, raimondo.manca@uniroma1.it

Credit ratings play nowadays a very important role in financial and risk management. Dynamics of credit ratings significantly influence investment flows at finance, insurance, energy, and other markets. At present, it becomes common to use Markov chains and related models to describe the dynamics of credit ratings as an indicator of the likelihood of rating default and other risk rating events. New stochastically ordered Markov and semi-Markov models for stochastic modelling of credit rating dynamics are introduced. Estimates for default time type distributions and their moment characteristics are presented. Simulation, numerical, and statistical aspects of using such models in credit rating calculations are also discussed.

Key words: Markov process, Semi-Markov process, Credit rating dynamics

Computational Modeling of Physico-mechanical, Electronic and Optical Properties of Nanoscale Materials (Nanotubes, Nanoparticle and Nanocomposites)

Katya Marinova Simeonova[1], Ganka Marinova Milanova[2]

Institute of Mechanics, Bulg. Acad. of Science
[1] Acad. G. Bonchev, str., Bl. 4, Sofia, 1113 Bulgaria, katyas@bas.bg
[2] UACG, Sofia, Bulgaria

Nanoscale materials possess extraordinary physico-mechanical properties (high elastiity moduli, high strentgh, high stiffness etc.), electronic, optical, electrical, transport, chemical, thermal and so on properties. Moreover these nanomatrials have very small nanosizes (about 50-100nm in tube' diameter) and chiral nanostructure. These both facts contibute very much to application of nanomaterials in technique, electronics, optoelectronics, transistors, nanoproducts, nanodevices, for space and energy applications. Recent articles, devoted to nanoscale materials improved that especially nanoparticles (nanorods, nanoshells), play an important role for cancer research in molecular and cellular biology, nanomedicine and biomedicine. Also, advamcements in novel nano and bio technologies are very actual ones, and could be considered as a revolution in the novel technologies. The aim of the work, proppoesed could be f! ormulated as follows: to give computatioal models for study of physico-mechanical behavior of carbon nanotubes (effects of the effective elastic characteristics, tube's geometry, effects of chirality). Computational modeling of nanorods and nanoshells' optical properties for cancer research technology, will be discussed as well. Nanocomposites as relatively novel materials (usally a polymer matrix, reinforced by carbon nanotubes), and thier mechanical, electroic, optical etc. properties will be modelled also. Numerical Fortran programs, dsigned by authors, have been used for numerical simulations. Graphics, reflecting different model's effects will be presented as well. Comparison, between experiments in literature and numerical results, obtained, show a very good agreement.
Key words: nanoscale materials, carbon nanotubes, nanorods, nanoshell, nanocomposites, computational modeling

Bayesian Analysis of Longitudinal Data Using Generalized Linear Mixed Models with the Same Marginal and Conditional Link

Debajyoti Sinha, Souparno Ghosh, Stuart Lipsitz and Bani Mallick

Dept of statistics, 117N. Woodward Ave, Tallahassee, Fl, USA, Sinhad@stat.fsu.edu

In this article, we present Bayesian methods of analysis for longitudinal outcomes using Generalized Linear Mixed models (GLMM) where the marginal link function, when integrated over the distribution of the random effects, is of the same form as that of the conditional link. We propose novel flexible models for random effects vector using a multivariate copula structure that allows a wide class of longitudinal association structures among outcomes from different time-points within a subject. Further, our Bayesian methods can accommodate the semiparametric models with non-parametric time-effects. We derive several important properties of our models, including proper posteriors of the fixed effects in the presence of non-informative priors. Our models and associated methodologies are illustrated with the analysis of longitudinal binary and count data from two AIDS studies.
Key Words: Bridge distribution, Copula model, partial linear model, probability integral transformation

Innovation Diffusion Modeling: An overview and applications

Christos H. Skiadas

Data Analysis and Forecasting Laboratory, Technical University of Crete
Chania, Crete, Greece, (e-mail: skiadas@otenet.gr)

Innovation diffusion modeling is a powerful tool for various applications in Sociology, Economics, Marketing, Management, Ecology and many other scientific fields. The related models are termed as innovation diffusion models but also as growth models and the associated growth functions expressing the growth or decline of a system.

The quantitative approaches to innovation diffusion modeling can be classified to 1) deterministic, 2) stochastic and 3) chaotic. Each of these three innovation diffusion modeling approaches is used to specific real life applications. Their use resulted in significant improvement of the theoretical and applied fields.

This review study explores the main quantitative approaches to innovation diffusion modeling and the progress in the related theory and applications.

Key words: Innovation diffusion, deterministic, stochastic, chaotic

The First Exit Time Theory and related applications to the modeling of Life Table Data sets

Christos H. Skiadas[1] and Charilaos Skiadas[2]

[1]Data Analysis and Forecasting Laboratory, Technical University of Crete
Chania, Crete, Greece, (e-mail: skiadas@otenet.gr)
Mathematics and Computer Science Department
Hanover College, Indiana, USA, (e-mail: Skiadas@hanover.edu)

The development and application of the first-passage-time theory for a stochastic process in modeling the human life table data is presented. Analytic derivation of the probability density function for the first exit time is given along with a review of the existing literature. The method of the tangent approximation to one-sided Brownian exit densities is preferred in order to find an analytic form of the first exit time densities. Some forms of the probability density functions are applied to mortality data and stochastic simulations are done.
Keywords: First exit time, Hitting time, Stochastic modeling, First-passage-time density, Health state function, Life table data, Stochastic simulation, The tangent approximation, One-sided Brownian exit densities.

Evaluation of Results in Discrete Discriminant Analysis

Ana Sousa Ferreira and Margarida Cardoso

LEAD, Faculty of Psychology and CEAUL, University of Lisbon, Portugal, asferreira@fp.ul.pt
UNIDE, Dep. of Quantitative Methods of ISCTE - Lisbon University Institute, Portugal,
margarida.cardoso@iscte.pt

Generally, the evaluation of results in Discrete Discriminant Analysis (DDA) is based on the data of a contingency table that associates the observed and predicted classifications (confusion matrix). It specifically focuses on the diagonal of the confusion matrix which expresses the number of elements in a priori classes that are classified in the same classes using the discriminant-classification rules.

In the present work, we suggest that the association between the a priori and the predicted classes can be further explored using indices of agreement between the observed and predicted partitions (e.g. (Cardoso (2007))). We illustrate our approach using simulated data (moderate and reduced size) with 2 and 4 classes and binary independent variables. We analyze the results of alternative classification methodologies: the well known CART algorithm (Breiman et al. (1984)) and a model (Brito et al. (2006)) that combines the Full Multinomial Model (FMM) and the First Order Independence Model (FOIM). Finally we discuss the obtained results trying to interpret the utility of each index in the context of DDA.
Key Words: CART, Combining models, Discrete Discriminant Analysis, Evaluation of results, Indices of agreement

References
1. Breiman, L., Friedman, J. H.; Olshen, R. A.; Stone, C. J., Classification and Regression Trees. California: Wadsworth, Inc. (1984).
2. Brito, I., Celeux, G. and Sousa Ferreira, A., Combining methods in supervised classification: A comparative study on discrete and continuous problems. REVSTAT - Statistical Journal, 4(3), 201-225 (2006).
3. Cardoso, M. G. M. S., Clustering and Cross-Validation. Paper presented at the IASC 07 - Statistics for Data Mining, Learning and Knowledge Extraction, Aveiro, Portugal (2007).

A Hybrid Genetic Algorithm for Solving the Uncapacitated Multiple Allocation Hub Location Problem

Zorica Stanimirović, Maja Djukić and Jozef Kratica

Faculty of Mathematics, University of Belgrade, Belgrade, Serbia, zoricast@matf.bg.ac.rs, mdjukic85@gmail.com
Mathematical Institute, Serbian Academy of Sciences and Arts, Belgrade, Serbia, jkratica@mi.sanu.ac.rs

In this paper, a hybrid genetic algorithm for solving the Uncapacitated Multiple Allocation Hub Location Problem is proposed. This NP-hard problem has significant application in designing modern transportation and telecommunication networks, such as road and railway systems, postal systems, systems of fast delivery, etc.
In order to improve the efficiency, genetic algorithm is combined with the local search heuristic. The proposed hybrid method shows to be very successful in solving problems of large dimensions with up to n=120 nodes. It is also tested on instances with n=130 and n=200 nodes for which no optimal solution is presented in the literature so far. Although the optimal solutions are not known, we believe that the proposed hybrid method provides high quality solutions on these problem instances unsolved to optimality before.
Keywords: Genetic Algorithms, Hub Location Problems, Combinatorial Optimization, Metaheuristics, Transportation and Telecommunication Networks.

Assessing the Effect of Informative Censoring in Piecewise Parametric Survival Models

Natalie Staplin

School of Mathematics, University of Southampton, UK, nds105@soton.ac.uk

Many of the standard techniques used to analyse censored survival data assume that there is independence between the failure time and censoring processes. There are situations where this assumption can be questioned, especially when looking at medical data. Wrongly assuming independence in a model could give misleading results. Therefore, it would be

useful to know whether we can assume independence or whether we need a model that takes account of any dependence. So we have developed a method that assesses the sensitivity of the parameter estimates in parametric models to changes in the assumed amount of dependence between failure time and censoring. We will account for the dependence in our model using a parameter δ and a bias function. We can think of δ as the correlation coefficient between the failure and censoring times. The sensitivity analysis is then conducted by varying the δ parameter by small amounts around zero, and estimating the change in the parameter of interest from that in the independent model. Parametric models with piecewise hazard functions are considered to allow a greater amount of flexibility in the models that may be fitted. In particular, piecewise exponential models will be fitted. This method is applied to a dataset that follows patients registered on the waiting list for a liver transplant. It suggests that in some cases even a small change in the amount of dependence could have a large effect on the results obtained.

Keywords: Survival analysis, informative censoring, sensitivity analysis, piecewise parametric models.

A Wavelet Based Prediction Method for Time Series

Cristina Stolojescu[1,2] Ion Railean[1,3] Sorin Moga[1] Philippe Lenca[1] and Alexandru Isar[2]

[1] Institut TELECOM; TELECOM Bretagne, UMR CNRS 3192, Lab-STICC; Universite europeenne de Bretagne, France, firstname.lastname@telecom-bretagne.eu
[2] Politehnica University of Timisoara, Romania, Faculty of Electronics and Telecommunications, firstname.lastname@etc.upt.ro
[3] Technical University of Cluj-Napoca, Romania, Faculty of Electronics, Telecommunications and Information Technology

The paper proposes a wavelet-based forecasting method for time series. We used the multi-resolution decomposition of the signal implemented using trous wavelet transform. We combined the Stationary Wavelet Transform (SWT) with four prediction methodologies: Artificial Neural Networks, ARIMA, Linear regression and Random walk. These techniques were applied to two types of real data
series: WiMAX network traffic and financial. We proved that the best results are obtained using ANN combined with the wavelet transform. Also, we compared the results using various types of mother wavelets. It is shown that Haar and Reverse biorthogonal 1 give the best results.

Keywords: time series, Stationary Wavelet Transform, forecasting.

Ranges, Limits and Boundaries of the Human Blood Pressures

Marijan Sviben

Zagreb, Croatia, marijan.sviben@zg.t-com.hr

High blood pressure may be a symptom of the essential hypertension - a disease of which 24-29% of elderly and more than 47% of old people suffer in Europe and North America. The cause of the disease is unknown. Peter and Richard Wingate (1988) have written that essential hypertension could not be properly defined because there is no clear division between the ranges of normal and abnormal pressures. Essential hypertension - Peter and Richard Wingate wrote - is divided into benign hypertension and the much rarer malignant hypertension. These terms - the authors continued - are necessarily vague, because nobody knows the limits of normal blood pressure. Sir George Pickering, however, in 1961 (WHO Report No. 862/1966) formulated the concept that blood pressure in a population is distributed continuously as a bell-shaped curve with no real separations between "normotension" and "hypertension" ... the dividing line between "normotension" and "hypertension" is arbitrary (and) can be defined only in an operational way. The operational classification of hypertension by blood pressure level could be done by the WHO Expert

Committee in 1994. Observing the only one human being from May 27th 1997 till February 19th 1998 resulted with the report that magnitudes of systolic blood pressure (SBP) and the diastolic blood pressure (DBP) as well as the pulse blood pressure (PBP) varied, fluctuating from a X minimum to a X maximum with a central tendency of expressing the character mostly around the mean value for the sample. The frequency distribution curves of magnitudes of blood pressures were alike to normal - Gaussian ones. The author continued his research work and he is presenting the frequency distributions of magnitudes of SBP and SBP during the hypotension, at a healthy middle-age (elder) person, at a healthy old human being as well as during the malignant hypertension. The ranges are visible. The limits of typical values are determined. The boundaries of normalcies and pathological conditions have been found out so that programme basis for the software evaluating the measurements of human blood pressures can be written.

The Moving of Magnitudes of the Blood Pressure of a Patient During His Recuperation After the Stroke

Marijan Sviben

Zagreb, Croatia, marijan.sviben@zg.t-com.hr

The moving of magnitudes of the blood pressures was observed at the man being 65 years of age who did not know that he suffered from the hypertension till he has gone through the stroke. The patient recovered with drugs, the rest and the special training. From May 27th till February 19th 1998 the magnitudes of the patient's systolic and diastolic blood pressure were measured 202 times with sphygmomanometer or with a kind of fully automatic blood pressure monitor (PIXEL or OMRON 3), regularly in the morning before the breakfast, at the end of acting of drugs taken during a day before and before the acting of drugs taken during a day of measuring. The intention was to establish whether the distributions of magnitudes of systolic, diastolic and the pulse blood pressure were like bell-shaped Gaussian curves and what the directions, the strength and the forms of the correlation (regression) between the magnitudes of systolic and pulse blood pressure were. The distributions of frequencies of magnitudes of systolic and the pulse blood pressure were like normal. The correlation between studied characteristics was positive, strong and very significant ($r_{x/y}$ = +0.631 > $r_{x/y}$ = 0.181 - d.f. = 200, P = 0.01). The form of correlation (regression) was expressed with the equation of trend II°:

$$Y_c = 58.73 + 3.025 X_c - 0.074 X_c^2.$$

(the origin: 161.5 mmHg as 0, the unit of X_c corresponds to 8 mm Hg)

The equation was used the magnitudes of the pulse blood pressure to be calculated and with them the magnitudes of systolic blood pressure to be diminished in order to find out the magnitudes of diastolic blood pressure expected at certain magnitudes of systolic blood pressure. The established values according to the magnitudes accepted by physicians generally health status of the middle aged persons or of the human beings older than 60 years of age to be estimated. During the research work it has been understood that the deviations of magnitudes of the pulse blood pressure from so called normal magnitudes of the pulse blood pressure are risky at the hypertension crisis.

FUZZY MARKOV SYSTEMS FOR THE DESCRIPTION OF OCCUPATIONAL CHOICES IN GREECE

M. A. Symeonaki[1] and R. N. Filopoulou[2]

Department of Social Policy, Panteion University of Social and Political Sciences, 136 Syggrou Av., 176 71, Athens, Greece.
[1] msimeon@panteion.gr , [2] celestfilopoulou@msn.com

In this paper the theory of fuzzy non-homogeneous Markov systems is applied for the first time to the study of occupational choices in Greece. This is an effort to deal with the

uncertainty introduced in the estimation of the transition probabilities and the difficulty of estimating their real values. In the case of studying the occupational choices of children, the traditional methods for estimating the probabilities cannot be used due to lack of data. The introduction of fuzzy logic into Markov systems provides us with a powerful tool, taking advantage of the heuristic knowledge that the experts of the system posses. The proposed model uses the symbolic knowledge of the occupational choices of children and focuses on the important factors that derive from the family environment and affect those choices. The aim is to develop a fuzzy expert system which best simulates the real conditions affecting the process of occupational choices in Greece.

Keywords: occupational choices, family factors, Markov systems, Fuzzy logic, Fuzzy Inference System

On deviations of the sample mean for Markov chains

Zbigniew S. Szewczak

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, ul. Chopina 12/18, 87-100 Toru´n, Poland, zssz@mat.uni.torun.pl

For homogeneous Markov chains satisfying the uniform recurrence condition Bahadur-Rao's expansions are established.

Keywords: large deviation, Markov chains.

MODELING RUMEN DEGRADATION OF FRESH 15N-LABELLED RYEGRASS AFTER ITS INGESTION BY SHEEP

R.Tahmasbi[1], J.V.Nolan1, R.C.Dobos[2]

[1] School of Rural Science and Agriculture, UNE, Armidale, NSW 2351;
[2] NSW Department of Primary Industries, Armidale, NSW 2351, Kerman, Iran
reza.tahmasbi@gmail.com

Quantitative models of N transactions in the rumen have been built using data obtained from in vivo experiments carried out using 15N dilution techniques. The Nitrogen (N) kinetics in the rumen were determined either by labelling the ruminal ammonia (NH3-N) pool (3) via intraruminal injection of 15NH4Cl or the sheep were allowed to ingest a single meal of 15N-labeled fresh ryegrass. The 15N enrichment was measured over time in ruminal NH3-N, bacterial N (2), peptide and amino acids (4) pools. Even, the same quantity of 15N was dosed into the rumen the NH3-N enrichment was always lower when sheep ingested fresh ryegrass. This finding indicates that 15N in ingested ryegrass was not completely degraded to ammonia in the rumen. It provides further evidence that production of ammonia is not an ideal indicator of fresh forage protein degradability in vivo. A model, similar to one described by Nolan (1975) for sheep given 800 g/d of chopped lucerne hay was modified to show the potential N transactions when fresh ryegrass protein is degraded in the rumen. The pool sizes for individual pools except for bacterial N and the flux rates between pools were obtained from in vivo experiment. The size of the bacterial pool used was taken from a study by Rodriguez et al.(2003). This model was used to simulate the enrichment v. time curves in Pools 2, 3 and 4 in response to an injection of 1 mmol of 15NH3 into Pool 3. Second, the model was used to simulate the enrichment curves when all the 15N ingested as ryegrass (1 mmol 15N) and released from Pool 1 over 1 day (assuming this occurred at an exponentially decreasing rate). In this case, the model did not satisfactorily simulate the enrichment v. time curves for Pools 2, 3 and 4 and the simulation of the 15N flowing through the rumen fluid peptide/amino acids pools was not sufficient to account for the rate of appearance of, or the actual level of,! labelling in microbial protein, i.e. the flux rate was much s! lower th an that found in vivo. It seemed necessary to postulate that another direct route of transfer of 15N from labelled ryegrass to microbial protein must exist from Pool 1 to Pool 4. The model was therefore modified and the arrow between ryegrass and the bacterial pool could represent microbes

attached to the ryegrass particles that assimilate some labelled peptides and amino acids directly from the plant, thus never entering the rumen fluid. This route would be separate from, and in addition to, any assimilation from the rumen fluid pool.
Key Words: Protein, metabolism, modeling

A Dynamical Recurrent Neuro-Fuzzy Algorithm for System Identification

Dimitris C.Theodoridis[1], Yiannis S. Boutalis[1], and Manolis A. Christodoulou[2]

[1] Democritus University of Thrace, Xanthi, Greece, 67100, dtheodo@ee.duth.gr
[2] Technical University of Crete, Chania, Crete, Greece, 73100, manolis@ece.tuc.gr

In this paper we discuss the identification problem which consists of choosing an appropriate identification model and adjusting its parameters according to some adaptive law, such that the response of the model to an input signal (or a class of input signals), approximates the response of the real system to the same input. For identification models we use fuzzy-recurrent high order neural networks. High order networks are expansions of the first order Hopfield and Cohen-Grossberg models that allow higher order interactions between neurons. In the present approach the HONN's used as approximators of the underlying fuzzy rules. New learning laws are proposed which ensure that the identification error converges to zero exponentially fast. There is a core idea in the proposed method: Several high order neural networks are specialized to work around fuzzy centers separating in this way the system in *neuro-fuzzy* subsystems which are associated with a number of fuzzy rules.
Keywords: Neuro-Fuzzy Systems, Identification, Gradient Descent, Pure Least Lquares.

A Probabilistic-Driven Search Algorithm for solving Multi-objective Optimization Problems

Thong Nguyen Huu, Hao Tran Van

Department of mathematics-Information, HCMC University of Pedagogy
280, An Duong Vuong, Ho Chi Minh city, Viet Nam
thong_nh2002@yahoo.com, tranvhao@gmail.com

This paper proposes a new probabilistic algorithm for solving multi-objective optimization problems, Probabilistic-Driven Search (PDS) algorithm. The algorithm uses probabilities to control the process in search of Pareto optimal solutions; specifically we use the absorbing Markov Chain to argue the convergence of the proposed algorithm. We test this approach by implementing the algorithm on some benchmark multi-objective optimization problems, and we find very good and stable results.
Keywords: Multi-objective, Optimization, Stochastic, Probability, Algorithm.

BERNSTEIN - VON MISES THEOREM IN BAYESIAN ANALYSIS OF COX MODEL

Jana Timková

Department of Statistics, Charles University, Prague
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic,
timkova@karlin.mff.cuni.cz

Although not well-known, the Bernstein-von Mises theorem (BvM) is a so-called bridge between bayesian and frequentist asymptotics. Basically, it states that under mild conditions the posterior distribution of the model parameter centered at the maximum likelohood estimator (MLE) is asymptotically equivalent to the sampling distribution of the MLE. This is a

powerful tool especially when the classical asymptotics is tedious or impossible to conduct while bayesian asymptotic properties can be obtianed via MCMC. However, in semiparametric setting with presence of infinite-dimensional parameters, as is e.g. Cox model for survival data, the results regarding BvM are more difficult to establish but still not impossible. The proposed poster gives short overview of BvM results found in the survival analysis context.

Bernstein-von Mises theorem and goodness-of-fit methods in Bayesian survival analysis

Jana Timková

Institute of Information Theory and Automation at Academy of Sciences of the Czech Republic, Prague
Department of Probability and Mathematical Statistics at Charles University, Prague
j.timkova@gmail.com

The contribution deals with analysis of event history type data. When we deal with survival data and the covariates are present, the famous Cox model is a convenient choice as being fairly easy to carry out the estimates of parameters. However, even though it is a semi-parametric model and therefore applicable to many various scenarios, it still has its restrictions. If we work with time-independent covariates then among first to assess is the proportional hazard assumption. Further, the model expects all individuals to behave according to the same baseline hazard function while in reality they might be stratified. Even if the model is properly chosen there may be need to check whether an important covariate has not been omitted or the functional form for influence of a covariate misspecified.

Goodness of fit testing based on martingale-residual processes (generally some transformation of the difference between observed and expected number of events) provides good way to detect potential defects in our model.

In the presented talk we shall focus our attention on the Bayesian approach to the Cox model where we will base our prior model on usage of Beta process as a prior for nonparametric part, i.e. cumulative hazard rate. Further, we will mention the recently appeared results connected with Bayesian asymptotics, as is the establishment of not very well-known Bernstein-von Mises theorem (BvM). The importance of this achievement lies in the fact that it represents a so-called bridge between Bayesian and frequentist asymptotics. Basically, it states that under mild conditions the posterior distribution of the model parameter centered at the maximum likelohood estimator (MLE) is asymptotically equivalent to the sampling distribution of the MLE. This is a powerful tool especially when the classical asymptotics is tedious or impossible to conduct while Bayesian asymptotic properties can be obtianed via MCMC. However, in semi-parametric setting with presence of infinite-dimensional parameters, as is e.g. Cox model for survival data, the results regarding BvM are more difficult to establish but still not impossible.

In the end, we will go through the goodness-of-fit methodologies based on residual processes and we will discuss the potential usage of them in the Bayesian concept.

 Keywords: Bayesian survival analysis, Cox model, infinite-dimensional parameters, Bernstein-von Mises theorem, goodness-of-fit, residuals

References
1. Andersen, P. K., Gill R. D., Cox's regression model for counting processes: A large sample study, *Ann. Statist.* **10**, 1100-1120 (1982).
2. Kim, Y., The Bernstein-von Mises theorem for the proportional hazard model, *Ann. Statist.* **34**, 1678-1700 (2006).
3. Marzec, L., Marzec, P., Generalized martingale-residual processes for goodness-of-fit inference in Cox's type regression models. *Ann. Statist.* **25**, 683-714 (1997).

Use and Misuse of the Weighted Freshét Distribution

George Tzavelas[1] and John, E. Goulionis[2]

Department of Statistics and Insurance Science, Universisty of Piraeus, 80 Karaoli & Dimitriou str. 185-35 Piraeus Greece.
[1]Email: tzafor@unipi.gr, [2]Email: sondi@otenet.gr

The Freshét distribution is considered as a proper model for the study of extreme phenomena. For the possibility of biased sampling, we consider the weighted Freshét distribution. Thus if $f(x, \beta, \sigma)$ is the pdf of a random variable *X*, which follows the Freshét distribution, the weighted Freshét distribution of order α is defined as

$$f_a(x, \beta, \sigma) = \frac{x^a}{\mu_a} f(x, \beta, \sigma)$$

where μ_a is the moment of order α. We develop the necessary equations for the maximum likelihood estimators, for the $f_a(x, \beta, \sigma)$ pdf, we prove its existence and some of its properties. We examine the misspecification effects on the various statistical features such as the mean, the variance and the median. With the term misspecification we mean to adopt the weighted Freshét distribution as the proper model for the analysis of the data while the correct model is the Freshét distribution. An alternative way of selecting the proper order α for the model $f_a(x, \beta, \sigma)$, is to treat α as an unknown parameter and to consider the maximum likelihood estimator of it. The theoretical results are applied to the 115-year annual maximum precipitation events for the Athens province.
Keywords: Freshét Distribution . Weighted Distribution, Maximum Likelihood Estimator.

Hidden Markov chain in the auditory system neural model

Yuriy V. Ushakov and Alexander A. Dubkov

Radiophysics faculty, University of Nizhny Novgorod, Nizhny Novgorod, Russia, nomah@list.ru
Radiophysics faculty, University of Nizhny Novgorod, Nizhny Novgorod, Russia, dubkov@rf.unn.ru

The noisy neural model of the auditory system of mammals is considered. The model is composed of two sensory neurons (*sensors*) receiving subthreshold sinusoidal input signals with different frequencies, and one output neuron (*interneuron*) receiving spike trains from sensors. All three neurons are also perturbed by uncorrelated white noise sources. Due to peculiarities of the model, its output spike train is non-Markovian. Nevertheless, the proposed probabilistic analysis reveals opportunity of model's behavior description by a Markov chain. We propose the calculation algorithm, which results in interspike intervals distribution curves approximating results of direct numerical simulations of the model with sufficient precision.
Keywords: Stochastic modeling, Stochastic differential equation, Leaky Integrate-and-Fire neuron, Interspike intervals, Hidden Markov chain.

References
1. Ushakov, Yu.V., Dubkov, A.A., and Spagnolo, B., The spike train statistics for consonant and dissonant musical accords, arXiv:0911.5243v1.
2. Ephraim, Y. Hidden Markov Processes, *IEEE transactions on information theory*, **48**(6), 1518 (2002).

Evaluation of M/G/1 Queues with Heavy-Tailed Service Time

Umay Uzunoglu Kocer

Department of Statistics, Dokuz Eylul University, Izmir, Türkiye, umay.uzunoglu@deu.edu.tr

The classical assumption of exponential service times has been problematical to satisfy in many real modern applications of queuing theory. Besides, recent evidence suggests that, heavy-tailed distributions have a significant role in some applications in computer network and telecommunications systems, and also some other applications in finance. We examine the performance of the M/G/1 queuing system when service time distribution is heavy-tailed. The analytical treatment of M/G/1 queues with heavy-tailed service time distribution is difficult. The queue performance is studied by using discrete event simulation and the effect of a long-tailed service time on the performance measures is examined.
Keywords: M /G /1 queue, service time distribution, simulation, long-tailed distributions

Branching process and Monte Carlo simulation for solving Fredholm integral equations

Kianoush Fathi Vajargah, Fatemeh Kamalzade and Farshid Mehrdoust

Department of statistics, Islamic Azad University, North Branch, Tehran, Iran
fathi_kia10@yahoo.com, fa.kamalzadeh@gmail.com
Department of Mathematics, University of Guilan, Rasht, Iran, fmehrdoust@guialn.ac.ir

In this paper we establish a new method for solving nonlinear Fredholm integral equations; however the type of one dimension nonlinear Fredholm integral equations were solved previously by Albert [1]. Now thinking about high dimension of integral equation is a cause of finding a marvelous relationship between branching process and Monte Carlo simulation, although this method require the optimum probability, integral simulation, and Monte Carlo algorithm.
Key words: Monte Carlo simulation, Markov chain, Fredholm integral equations, Branching process

Smoothing Parameter Selection in Hazard Rate Estimation for Randomly Right Censored Data

Francois C. van Graan

North-West University, Potchefstroom Campus, South Africa, Francois.VanGraan@nwu.ac.za

In this paper two methods for smoothing parameter selection in hazard rate estimation are considered. Firstly, a bootstrap method based on an asymptotic representation of the mean weighted integrated squared error is considered. Secondly, the least squares cross-validation method is considered. Both methods are compared by means of an empirical study and suggestions on how to improve the methods are given.
Keywords: Hazard rate, Smoothing parameter, Bootstrap method, Cross-validation, Censored data

References
1. Cao, R., Conzales-Manteiga W. and Marron, J.S., Bootstrap selection of the Smoothing parameter in Nonparametric Hazard Rate Estimation, *Journal of the American Statistical Association.,* **91,** 1130-1140(1996).

2. Patil, P.N., On the Least Squares Cross-Validation Bandwidth in Hazard Rate Estimation, *The Annals of Statistics,* **21,** 1792-1810(2003).

3. Tanner, M.A., and Wong, W.H., Data-Based Nonparametric Estimation of the Hazard Function With Applications to Model Diagnostics and Exploratory Analysis, *Journal of the American Statistical Association.,***79,** 174-182 (1984).

Composite likelihood methods

Cristiano Varin, Nancy Reid and David Firth

Department of Statistics, Ca' Foscari University, San Giobbe, Cannaregio, 873, Venice, 30121 Venice, Italy, sammy@unive.it

Composite likelihoods are pseudolikelihoods constructed by compounding marginal or conditional densities. In various applications, composite likelihoods are convenient surrogates for the ordinary likelihood when it is too cumbersome or impractical to be computed. This talk provides a survey of recent developments in the theory and application of composite likelihood. A range of application areas, including geostatistics, spatial extremes and space-time models as well as clustered and longitudinal data and time series are considered. Emphasis is given to the development of the theory, and the current state of knowledge on efficiency and robustness of composite likelihood inference.

Key Words: geostatistics, Godambe information, longitudinal data, pseudo-likelihood, robustness, spatial extremes, time series

Performance Analysis of an ant-based clustering algorithm

Rosangela Villwock, Maria Teresinha Arns Steiner and Pedro José Steiner Neto

UFPR, Curitiba, Pr, Brasil, pedrosteiner@ufpr.br

The collective behaviors and self-organization of social insects have inspired researchers to reproduce this behavior. Methods inspired in ants are a great promise for clustering problems. The main objective of the present paper was to evaluate the performance of the Clustering Modified Algorithm based in Ant Colony in relation to other modification of the algorithm denominated ACAM. The main changes were: substitution of the picked pattern by the ant in case it has not been dropped in 100 consecutives iterations; comparison of the probability of dropping a pattern in a certain position with the probability of dropping this pattern in the current position; evaluation of the probability of dropping a pattern in a new position in case the pattern has not been dropped in the position of raffle draw, but in a neighboring position. For the evaluation of the algorithm performance, two real databases were used. The results showed that the algorithm proposed in this work was better than the ACAM for the two databases.

Key Words: Metaheuristics; Data Mining; Clustering

On Energy Based Cluster Stability Criterion

Zeev Volkovich, Dvora Toledano Kitai and Renata Avros

Software Engineering Department, ORT Braude College of Engineering, Karmiel, Israel
vlvolkov@braude.ac.il, dvora@braude.ac.il, r_avros@braude.ac.il

In the current article, we purpose a method for the study of cluster stability. This method adopts a physical point of view. Such standpoint suggests using a physical magnitude of samples mixing within clusters constructed by means of a clustering algorithm. We quantify samples closeness by the relative potential energy between items belonging to different samples for each one of the clusters. This potential energy is closely linked with a '' gravity"

force between the two samples. If the samples, within each cluster, are well mingled, this quantity is sufficiently small. As known from electrostatics, if the sizes of the samples grow to infinity then the total potential energy of the pooled samples calculated for a potential, tends to zero in the case when the samples are drawn from the same population. The two-sample energy test has been constructed by G. Zech and B. Aslan based upon this perception. The statistic of the test measures the potential energy of the combined samples. Actually, we use this function as a characteristic of clustered samples similarity. The partition merit is represented by the worst cluster corresponding to the maximal potential energy value. To ensure readiness of the proposed model and to decrease the uncertainty in the model we draw many pairs of samples for a given number of clusters and construct an empirical distribution of the potential energy corresponding to the partitions created within the samples. Among all those distributions, one can expect that the true number of clusters can be characterized by the empirical distribution which is most concentrated at the origin. Numerical experiments, provided by means of the proposed methodology, demonstrate high ability of the approach.

Keywords: Clustering, Cluster stability, Two-sample energy test

Non-Standard Behavior of Density Estimators for Functions of Independent Observations

Wolfgang Wefelmeyer

Mathematical Institute, University of Cologne, 50931 Cologne, Germany, wefelm@math.uni-koeln.de

Densities of functions of two or more independent random variables can be estimated by local U-statistics. Frees (1994) gives conditions under which they converge pointwise at the parametric root-n rate.

Gin\'e and Mason (2007) give conditions under which this rate also holds in L_p-norms. We present several natural applications in which the parametric rate fails to hold in L_p or even pointwise.

1. The density estimator of a sum of squares of independent observations typically slows down by a logarithmic factor. For exponents greater than two the estimator behaves like a classical density estimator.

2. The density estimator of a product of two independent observations typically has the root-n rate pointwise, but not in L_p -norms.

An application is given to semi-Markov processes and estimation of an inter-arrival density that depends multiplicatively on the jump size.

3. The stationary density of a nonlinear or nonparametric autoregressive time series driven by independent innovations can be estimated by a local U-statistic (now based on dependent observations and involving additional parameters), but the root-n rate can fail if the derivative of the autoregression function vanishes at some point.

This is joint work with Anton Schick (Binghamton University).

Keywords: density estimator, convergence rate, U-statistic.

Fund-of-Funds Construction by Statistical Multiple Testing Methods

Michael Wolf and Dan Wunderli

University of Zurich, Zurich, Switzerland, mwolf@iew.uzh.ch

Fund-of-funds (FoF) managers face the task of selecting a (relatively) small number of hedge funds from a large universe of candidate funds. We analyse whether such a selection can be successfully achieved by looking at the track records of the available funds alone, using advanced statistical techniques. In particular, at a given point in time, we determine which funds significantly outperform a given benchmark while, crucially, accounting for the fact that a large number of funds are examined at the same time. This is achieved by employing so-

called multiple testing methods. Then, the equal-weighted or the global minimum variance portfolio of the outperforming funds is held for one year, after which the selection process is repeated. When backtesting this strategy on two particular hedge fund universes, we find that the resulting FoF portfolios have attractive return properties compared to the $1/N$~portfolio (that is, simply equal-weighting all the available funds) but also when compared to two investable hedge fund indices.

Key Words: Bootstrap, familywise error rate, fund-of-funds, performance evaluation

Modelling covariance structure for multivariate longitudinal data

Jing Xu[1] and Gilbert MacKenzie[1,2]

[1]Centre of Biostatistics, University of Limerick, Ireland, jing.xu@ul.ie
[2]ENSAI, Rennes, France, gilbert.mackenzie@ul.ie

In many epidemiological studies and clinical trials, subjects are measured on several occasions with regard to a collection of response variables. Analysis of such multivariate longitudinal data involves modelling the joint evolution of the response variables over time. Consider, as an example, The San Luis Valley Diabetes Study (SLVDS) (Jones 1993). LSVDS is a case control study of non-insulin-dependent diabetes mellitus (NIDDM) in the San Luis Valley in southern Colorado. The 173 subjects had between one and four visits to the clinic. For them, two response variables, body mass index (BMI) and the subject's age at the time of visit, are recorded. There are many similar examples (Chapman et al 2003, Thiebaut et al 2002, Newsom 2002, and McMullan, *et al*, 2003).

However, the analysis of such multivariate longitudinal data is complicated by the existence of correlation between the responses at each time point, the correlation within separate responses over time and the cross-correlation between different responses at different tines. Therefore, one major task in analyzing these data is to model efficiently the covariance matrices $cov(y_i)=\Sigma_i$ for $i=1,\ldots,n$ subjects. Several approaches have been developed: doubly multivariate models (DMM) analysis (Timm, 1980), multivariate repeated measurement models with a Kronecker product covariance structure (Galecki, 1994), multivariate mixed models (Jones 1993) and structural equation modelling approach (Hatcher, 1998).

In this paper, we develop a data-driven method to model the covariance structures. For simplicity, consider a repeated measures study in the bivariate case. Let $y_{ij} = (y_{ij}^1, y_{ij}^2)$ represent the observation of two response variables for the i-th individual at j-th time point ($i=1,\ldots,n; j=1,\ldots,m$). Further set $y_i = (y'_{i1},\ldots, y'_{im})'$. The covariance matrices of each y_i is denoted by Σ_i which has $\Sigma_{jk} = E(y_{ij} y'_{ij})$ as its (j,k)-th sub-matrix. Noting that Σ_i is positive definite it can be factorised in block triangles (Hamiliton, 1994).

$$T_i \Sigma_i T_i' = D_i$$

where. T_i is a block triangular matrix with 2 x 2 identity matrices as diagonal entries and D_i is a block-diagonal matrix with positive definite 2 × 2 matrices as diagonal entries. Furthermore, the block triangular factorization of Σ_i has a nice statistical interpretation from the perspective of linear prediction in time series. We refer to the new parameters, Θ_{ijk}'s, the block below-diagonal entries of T_i, and the D_{ij}s, the block diagonal entries of D_i, as the autoregressive coefficient matrices and the innovation matrices of Σ_i. After taking matrix logarithm of D_{ij}, the $2m(2m+1)/2$ constrained and hard-to-model parameters of Σ_i can be traded in for the $2m(2m+1)/2$ unconstrained and interpretable parameters, Θ_{ijk}. We show how estimates of these parameters, and the parameters in the mean, can be obtained by Maximum Likelihood with Normally distributed responses and how the optimal combination of the three sets of parameters may be chosen using AIC and BIC criteria.

Branching Walks in Homogeneous and Inhomogeneous Random Environments

Elena Yarovaya

Dept. Probability Theory, Faculty of Mechanics and Mathematics, Moscow State University, Moscow, Russia, yarovaya@mech.math.msu.su

We consider models of a continuous-time branching random walk on the multidimensional lattice in a random environment. In the first of these models, the branching medium formed of birth-and-death processes at every lattice site is assumed to be spatially homogeneous. In the second one the branching medium containing a single source of branching situated at the origin is spatially inhomogeneous. In the both models the underlying random walk is assumed to be symmetric. We study the long-time behavior of the moments for the local and total particle populations, obtained by averaging over the medium realizations. Conditions at which the long-time behavior of the moments for the numbers of particles at an arbitrary site of the lattice and on the entire lattice coincides for the both models are obtained. It has been shown that these conditions hold for the random potential with a Weibull type upper tail or a double exponential type upper tail. The results obtained are useful for planning of experiments in the frame of these models.
Key Words: Branching random walks, Inhomogeneous random environment, Homogeneous random environment, Kolmogorov backward equations, Moments, Intermittency.

Application of irregular sampling in statistics of stochastic processes

Vladimir Zaiats

Escola Politècnica Superior, Universitat de Vic, c/. Laura, 13, Vic (Barcelona), Catalonia, Spain, vladimir.zaiats@uvic.es

There is a wide range of methods for estimation of functional characteristics (correlation functions, spectral densities, etc.) of stochastic processes. These methods are based either on frequency- or on time-domain methods. When working in the time domain, the model if often parameterized by means of a finite number of parameters. In practice, it may not be reasonable to assume that the "true" model admits a finite parameterization. In this context, the purpose of system modeling is to obtain a model involving a finite number of unknown parameters that provides a "reasonable" approximation to the observed data, rather than to estimate parameters of the "true" system.

We would like to discuss the existing methods of estimation of correlation functions, spectral densities, impulse transfer functions specially focusing on the possibility of application of irregular sampling which has proves to be efficient in estimation of correlation functions and spectral densities. Some new results in non-parametric estimation of the transfer function will be presented. We will also discuss possible extensions to the framework of random fields with recent applications to seismological and cosmological data.

References

1. Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D., Asymptotics for spherical needlets, *Ann. Statist.*, **37**, 1150-1171 (2009).
2. Buldygin, V., Utzet, F. and Zaiats, V., Asymptotic normality of cross-correlogram estimates of the response function, *Stat. Inference Stoch. Process.*, **7**, 1-13 (2004)
3. Goldenshluger, A., Nonparametric estimation of transfer functions : rated of converge nce and adaptation, *IEEE Trans. Inform. Theory*, **44**, 644-658 (1998).

4. Lii, K.-S. and Masry, E., Model fitting for continuous time stationary processes from discrete-time data, *J. Multivar. Anal.*, **41**, 56–79 (1992).
5. Masry. E., Nonparametric regression estimation for dependent functional data: asymptotic normality, *Stochastic Process. Appl.,* **115**, 155-177 (2005).
6. Rachdi, M., Strong consistency with rates of spectral estimation of continuous-time processes: from periodic and poisson sampling schemes, *J. Nonparametr. Statist.*, **16**, 349-364 (2994).

On Asymptotic Comparison of Maximum Likelihood Estimators of the Shape Parameter

Alexander Zaigraev

Nicolaus Copernicus University, Toruń, Poland, alzaig@mat.uni.torun.pl

Let a sample drawn from the shape-and-scale family of distributions be given. We are interested in the estimation of the shape parameter while the scale parameter is considered as a nuisance parameter.

Two estimators of the shape parameter are used. The first one is the usual maximum likelihood estimator. This estimator is scale invariant. The second one is the maximum likelihood scale invariant estimator (see Zaigraev and Podraza-Karakulska, 2008a, b, where the case of gamma distribution was considered). It is obtained by applying the maximum likelihood principle to the measure defined on the σ–algebra of scale invariant sets generated by the underlying distribution (see, e.g. Hájek et al., 1999, Subsection 3.2.2). As a measure of comparison of those two estimators the mean square error is used.

We consider the asymptotic case and with the help of undefined factors we obtain the expansions for both estimators, as well as for their mean square errors. It is shown that the maximum likelihood estimator is worse than the maximum likelihood scale invariant estimator. As examples, we consider the cases of gamma distributions and Weibull distributions. The asymptotic expansions for both estimators, as well as the asymptotic expansions for their mean squared errors, are obtained. The results are supported with numerical calculations.

Keywords: Shape-and-scale family, Maximum likelihood estimator, Maximum likelihood scale invariant estimator, Mean square error, Undefined factors, Gamma distribution, Weibull distribution.

References
1. Hájek, J., Šidák, Z. and Sen, P. K., *Theory of Rank Tests*. Second ed. Academic Press, San Diego (1999).
2. Zaigraev, A. and Podraza-Karakulska, A., On estimation of the shape parameter of the gamma distribution. *Statist. Probab. Letters.* **78** (2008), 286--295.
3. Zaigraev, A. and Podraza-Karakulska, A., On asymptotics of the maximum likelihood scale invariant estimator of the shape parameter of the gamma distribution. *Appl. Math. (Warsaw).* **35** (2008), 33--47.

Estimates for the Rate of Strong Approximation in the Multidimensional Invariance Principle

Zaitsev Andrei Yu.

St.-Petersburg Department of Steklov Mathematical Institute, Fontanka 27, 191023, Russia
zaitsev@pdmi.ras.ru

We consider the problem of constructing on a probability space a sequence of independent **R**^{d}-valued random vectors $X1, \ldots, Xn$ (with given distributions) and a corresponding sequence of independent Gaussian random vectors $Y1, \ldots, Yn$ so that the quantity

$$\Delta(X,Y) = \max_{1 \le k \le n} \left\| \sum_{j=1}^{k} X_j - \sum_{j=1}^{n} Y_j \right\|$$

would be so small as possible with large probability. The estimation of the rate of strong approximation in the invariance principle may be reduced to this problem.

Another problem is to construct infinite sequences of i.i.d $X1, \ldots ,Xn, \ldots$ (with given distributions) and a corresponding sequence of i.i.d. Gaussian random vectors $Y1, \ldots , Yn \ldots$ so that $\left\| \sum_{j=1}^{k} X_j - \sum_{j=1}^{n} Y_j \right\| = O(f(n))$ almost surely with slowly increasing $f(n)$.

We formulate the results published in the papers of Zaitsev [6], [7] and Gotze and Zaitsev [2], [3]. They may be considered as multidimensional generalizations and improvements of some results of Komlos, Major and Tusnady [4], Sakhanenko [5] and Einmahl [1].

References
1. Einmahl, U., "Extensions of Results of Komlos, Major and Tusnady to the Multivariate Case". J. Multivar. Anal. 28:20–68 (1989).
2. Gotze, F. and Zaitsev, A. Yu., "Bounds for the Rate of Strong Approximation in the Multidimensional Invariance Principle". Theor. Probab. Appl. 53:100–123 (2008).
3. Gotze, F. and Zaitsev, A. Yu., "Rates of Approximation in the Multidimensional Invariance Principle for Sums of i.i.d. Random Vectors with Finite Moments". Zapiski Nauchnykh Seminarov POMI 368:110–121 (2009).
4. Komlos, J., Major, P., and Tusnady, G., "An Approximation of Partial Sums of Independent RV'-s and the Sample DF. I; II", Z. Wahrscheinlichkeitstheor. verw. Geb. 32, 111–131 (1975); 34, 34–58 (1976).
5. Sakhanenko, A. I., "Estimates in the Invariance Principles", in Trudy Inst. Mat. SO AN SSSR, v.5, Nauka, Novosibirsk, 27–44 (1985).
6. Zaitsev, A. Yu., "Estimates for the Rate of Strong Gausian Approximation for Sums of i.i.d. Random Vectors". Zapiski Nauchnykh Seminarov POMI 351:141–157 (2007).
7. Zaitsev, A. Yu., "Rate for the Strong Approximation in the Multidimensional Invariance Principle". Zapiski Nauchnykh Seminarov POMI 364:148–165 (2009).

Title Index

111

114

Author Index

Gromov, Aleksander	20	Kano, Hiroyuki	41
Guastaldi, Enrico	45	Karagrigoriou, Alex	51
Guerbyenne, Hafida	46	Khaldi, Khaled	51
Guermeur, Yann	46	Kianfar, Farhad	80
Guerrizio, Maria Alessandra	10	Kim, Sahm	52
		Kimber, Alan	52
Guigou, Jean-Daniel	46	Kitsos, Christos P	53
Guillen, Montserrat	29	Kokolakis, George	54
Guolo, Annamaria	44	Koole, Ger	47
Guyon, X.	47	Kort, Peter M.	24
Haensel, Alwin	47	Kosolapov, Samuel	53
Haim, Klaus Dieter	64	Athanasios Kottas	55
Hamdi, Fayçal	46	Kouvaras, George	54
Hamed, Shakouri G.,	69	Kovalenko, Anatoly	11
Hamilton-Keene, Rachael	20	Kratica, Jozef	95
		Krit, Makram	54
Hamouda, Mohamed Ben	64, 65	Krnjajić, Milovan	55
Hardouin, Cécile	47	Kryukov, Aleksey K.	77
Hartl, Richard F.	24	Kurilovas, Eugenijus	55
Hartney, Philippa	20	Laniado, Henry	55
Helman, Karel	47	Lantuéjoul, C.	14
Hernandez-del-Valle, Gerardo	48	Laouar, A.	3
		Larramendy-Valverde, Irène	50
Hinde, John	27, 48	Lavergne, Christian	87
Hu, Ying	71	Le Calvé, Georges	33
Hunt, Julien S.	49	Leandre, Remi	56
Húsek, Dušan	83	Lefèvre, Claude	42, 57
Hutchings, MR	31	Lemeshko, Boris Yu	57
Huu, Thong Nguyen	99	Lemeshko, Stanislav B.	57
Ignaccolo, Rosaria	49	Lenca, Philippe	96
Ioannou, Petros	49	Li, Junping	58
Isar, Alexandru	96	Lillo, Rosa E.	40, 55
Janghorbani, M.	91	Lindqvist, Bo Henry	58
Jannsen, Jacques	50	Lipinski, Piotr	58
Jimenez, Javier	62	Lipsitz, Stuart	93
Joannides, Marc	50	Litvinenko, Alexander	59
Kaiser, Regina	14	Loubes, Jean-Michel	28
Kalamatianou,	15	Lovat, Bruno	46
Kamalzade, Fatemeh	102	Lundgren, Robin	92
Kanniainen, Juho	50		