

PROGETTO DI UNITÀ DI RICERCA - MODELLO B
Anno 2007 - prot. 20077P5AWA_002

1 - Area Scientifico-disciplinare

13: Scienze economiche e statistiche 100%

2 - Durata del Progetto di Ricerca

24 Mesi

3 - Coordinatore Scientifico

BATTAGLIA FRANCESCO

Professore Ordinario

Università degli Studi di ROMA "La Sapienza"

Facoltà di SCIENZE STATISTICHE

Dipartimento di STATISTICA, PROBABILITÀ E STATISTICHE APPLICATE

4 - Responsabile dell'Unità di Ricerca

GIANNERINI

SIMONE

Ricercatore non confermato

17/12/1970

GNNSMN70T17C296V

Università degli Studi di BOLOGNA

Dipartimento di SCIENZE STATISTICHE

0512098262
(Prefisso e telefono)

051232153
(Numero fax)

simone.giannerini@unibo.it

5 - Curriculum scientifico

Testo italiano

POSIZIONE ATTUALE:

Ricercatore, Università di Bologna, Facoltà di Economia, Rimini.
Titolare dei Corsi di "Statistica Applicata" e "Statistica per le decisioni di impresa".
Affersce al Dipartimento di Scienze Statistiche, via delle belle arti 41, 40126 Bologna.

FORMAZIONE

1998 - Laurea in Scienze Statistiche Demografiche e Sociali, Università degli studi di Bologna, tesi in Statistica applicata alle scienze fisiche: "Test statistici per fenomeni caotici"
Relatore: Rodolfo Rosa;
Correlatore: Antonella Capitanio;

2000 - research student alla London School of Economics

2001 - Master in Statistica (MSc Statistics) alla London School of Economics.

2002 - Dottorato di ricerca in "Metodologia statistica per la ricerca scientifica" presso l'Università degli Studi di Bologna, XIV ciclo. Tesi: "Sensitive dependence on initial conditions: chaos and stochastic processes".
Supervisor: Domenico Costantini, Howell Tong

2001-2003 Titolare dell'assegno di ricerca:
"Metodi sequenziali in ricerche epidemiologiche: schema della rilevazione sperimentale e articolazione delle procedure inferenziali"
Responsabile: Alessandra Giovagnoli.

2004 Titolare dell'assegno di ricerca:
"Trasformazioni di Serie Temporali: l'impatto sulla Stima e sulla Previsione"
Responsabile: Estela Bee Dagum.

RICERCA

Metodologia

-- Analisi delle serie storiche, teoria del caos e Statistica, Processi stocastici, Statistica computazionale, metodi di ricampionamento per dati dipendenti.

Applicazioni

-- Bioinformatica, Epidemiologia, Trial clinici, Analisi di flussi turistici, Economia dell'arte,

ha servito da referente per le seguenti riviste scientifiche:

Chaos, Econometric Reviews, Studies in Nonlinear Dynamics and Econometrics, Statistica, SMA.

Ha collaborato attivamente all'Unità di Bologna coordinata dalla prof. Estela Bee Dagum (coordinatore nazionale) nei Progetti di ricerca finanziati dal Miur negli anni 2000, "Linearità e non linearità nelle dinamiche di serie storiche", 2002, "Inferenza statistica sulla dinamica stocastica e deterministica di serie storiche osservate" e 2004, "Stima parametrica e non parametrica e previsione delle dinamiche di momenti condizionali di serie storiche".

E' titolare del progetto tseriesChaos una libreria R per l'analisi delle serie storiche non lineari motivate dalla teoria del caos.

APPARTENENZA AD ASSOCIAZIONI

Socio ordinario SIS dal 1999 (Società Italiana di Statistica).

Socio ordinario RSS dal 2002 (Royal Statistical Society).

Socio ordinario ASA dal 2004 (American Statistical Association).

Socio ordinario SICCC dal 1999 (Società Italiana Caos e Complessità).

Socio vitalizio ADI dal 1999 (Associazione Dottori di ricerca Italiani).

Testo inglese

POSITION:

Assistant Professor, University of Bologna, Faculty of Economics, Rimini.
Statistics Department, University of Bologna.

TEACHING

"Statistica Applicata" and "Statistica per le decisioni di impresa".
Faculty of Economics, Rimini

FORMATION

1998 - Degree in Statistics, Università degli studi di Bologna, in Statistics for Physics:

2000 - Research student at London School of Economics

2001 - MSc Statistics at London School of Economics.

2002 - PhD in Statistics, Università degli Studi di Bologna, XIV ciclo. Thesis: "Sensitive dependence on initial conditions: chaos and stochastic processes".
Supervisor: Domenico Costantini, Howell Tong

2001-2003 Research fellow with grant in
"Sequential methods in Epidemiology", Scientific advisor: Alessandra Giovagnoli.

2004 Research fellow with grant in
"Time series transformations: the impact on estimation and prediction"
Scientific advisor: Estela Bee Dagum.

RESEARCH

Metodology

-- Time series analysis, Chaos theory and Statistics, Stochastic processes, Computational Statistics, Resampling methods for dependent data.

Applications

-- Bioinformatics, Epidemiology, applied Economics.

Has served as referee for the following journals:

Chaos, Econometric Reviews, Studies in Nonlinear Dynamics and Econometrics, Statistica, SMA.

Served as an active collaborator to the Research Unit of Bologna (unit and national coordinator, prof. Estela Bee Dagum) to the Miur projects granted in 2000, "Linearity and non linearity in time series dynamics", 2002, "Statistical inference on the stochastic and deterministic dynamics of observed time series" e 2004, "Parametric and non parametric estimation and forecasting of time series conditional moment dynamics".

He is author of the library tseriesChaos a R package for time series analysis motivated by chaos theory.

Membership

Fellow SIS from 1999 (Italian Statistical Society).

Fellow RSS from 2002 (Royal Statistical Society).

Fellow SICCC from 1999 (Society for Chaos and Complexity).

Fellow ASA from 2004 (American Statistical Association).

Fellow ADI from 1999 (PhD students' Association).

6 - Pubblicazioni scientifiche più significative del Responsabile dell'Unità di Ricerca

1. A. BALDI ANTOGNINI, GIANNERINI S. (2007). Generalized Pòlya Urn designs with null balance. *JOURNAL OF APPLIED PROBABILITY*. vol. 44 (3), pp. 661 - 669 ISSN: 0021-9002. [ARTICOLO].
2. G. CANDELA, GIANNERINI S., A. E. SCORCU. (2007). Flussi turistici e caratteristiche strutturali dei turismi e delle destinazioni: una nota introduttiva. *ECONOMIA DEI SERVIZI*. vol. 1, Anno II, pp. 47 - 57 ISSN: 1970-4860. [ARTICOLO].
3. G. CANDELA, GIANNERINI S., A. E. SCORCU. (2007). Le caratteristiche strutturali di una destinazione balneare matura. Il caso di Rimini. *ECONOMIA DEI SERVIZI*. vol. 1, Anno II, pp. 123 - 146 ISSN: 1970-4860. [ARTICOLO].
4. GIANNERINI S., D. L. GONZALEZ, R. ROSA. (2007). Testing Chaotic dynamics in systems with two positive Lyapunov exponents: a bootstrap solution. *INTERNATIONAL JOURNAL OF BIFURCATION AND CHAOS IN APPLIED SCIENCES AND ENGINEERING*. vol. 17, n°1, pp. 169 - 182 ISSN: 0218-1274. [ARTICOLO].
5. GIANNERINI S., E. MAASOUMI, E. BEE DAGUM. (2007). An entropy based test for non-linear dependence in time series. *S.Co. 2007, Complex models and computational intensive methods for estimation and prediction*. 06-08 settembre 2007. (pp. 249 - 253). ISBN/ISSN: 978-88-6129-114-0. [RELAZIONE]. PADOVA: CLEUP (ITALY).
6. GIANNERINI S., E. MAASOUMI, E. BEE DAGUM. (2007). Entropy testing for nonlinearity in time series. *56th session of the International Statistical Institute*. 22-29 agosto 2007. [RELAZIONE]. LISBOA: International Statistical Institute (PORTUGAL).
7. A. DESALVO, GIANNERINI S., R. ROSA. (2006). Chaotic phenomena of charged particles in crystal lattices. *CHAOS*. vol. 16(2) 023114, pp. 1 - 12 ISSN: 1054-1500. [ARTICOLO] Impact Factor 2004: 1.942, Ranked 3/162 MATHEMATICS, APPLIED; 3/34 PHYSICS, MATHEMATICAL.
8. BALDI ANTOGNINI A., GIANNERINI S. (2006). Convergence rate for Ehrenfest-type urn designs. *XLIII Riunione Scientifica della Società Italiana di Statistica*. 14-16 giugno 2006. (pp. 495 - 498). [RELAZIONE]. s.l: s.n (ITALY).
9. C. BONETTO, GIANNERINI S., A. GIOVAGNOLI. (2006). The analysis of contingency tables with ordinal data: an application to monitoring antibiotic resistance. *STATISTICS IN MEDICINE*. vol. 25, pp. 3560 - 3575 ISSN: 0277-6715. [ARTICOLO].
10. D.L. GONZALEZ, GIANNERINI S., R. ROSA. (2006). Detecting Structure in Parity Binary Sequences. *IEEE ENGINEERING IN MEDICINE AND BIOLOGY MAGAZINE*. vol. 25, n°1, pp. 69 - 81 ISSN: 0739-5175. [ARTICOLO] Impact Factor 2005: 1.232.
11. E. BEE DAGUM, GIANNERINI S. (2006). A critical investigation on detrending procedures for non-linear processes. *JOURNAL OF MACROECONOMICS*. vol. 28 (1), pp. 175 - 191 ISSN: 0164-0704. [ARTICOLO] Impact Factor 2005: 0.177.
12. A. BONIFACINO, V. PETROCELLI, T. PISANI, GIANNERINI S., GIOVAGNOLI A., A. VECCHIONE, P.L. MINGAZZINI, M.R. GIOVAGNOLI. (2005). Accuracy Rates of US-guided Vacuum-assisted Breast Biopsy. *ANTICANCER RESEARCH*. vol. 25(4), pp. 2465 - 2470 ISSN: 0250-7005. [ARTICOLO].
13. A. DESALVO, GIANNERINI S., R. ROSA. (2004). Chaotic Phenomena Arising in the Interaction of MeV Protons with Silicon Crystals. *Workshop on Nonlinear Dynamics and Complexity in Information and Communication Technology*. 6-8 Settembre 2004. [RELAZIONE]. BOLOGNA: s.n (ITALY).
14. GIANNERINI S. (2004). Deterministic and Stochastic Linear and Nonlinear Mean Prediction of Time Series. *ASA Joint Statistical Meeting*. 8-12 Agosto 2004. [RELAZIONE]. ALEXANDRIA, VA: ASA (UNITED STATES).
15. GIANNERINI S., D.L. GONZALEZ, R. ROSA. (2004). A new Insight into DNA: detecting structure in parity binary sequences. *ECC8, Experimental Chaos Conference*. 14 Giugno 2004. [RELAZIONE]. FIRENZE: s.n (ITALY).
16. GIANNERINI S., R. ROSA. (2004). Assessing chaos in time series: statistical aspects and perspectives. *STUDIES IN NONLINEAR DYNAMICS AND ECONOMETRICS*. vol. 8, Article 11 ISSN: 1081-1826. [ARTICOLO] Impact Factor 2004: 0.242.

7 - Elenco dei partecipanti all'Unità di Ricerca

7.1 - Componenti

Componenti della sede dell'Unità di Ricerca

n°	Cognome	Nome	Università/Ente	Qualifica	Impegno	
					1° anno	2° anno
1.	GIANNERINI	Simone	Università degli Studi di BOLOGNA	Ricercatore non confermato	6	6
2.	GUIDOTTI	Laura	Università degli Studi di BOLOGNA	Professore Associato confermato	6	6
3.	LUATI	Alessandra	Università degli Studi di BOLOGNA	Ricercatore confermato	6	6
TOTALE					18	18

Componenti di altre Università / Enti vigilati

Nessuno

Titolari di assegni di ricerca

Nessuno

Titolari di borse

n°	Cognome	Nome	Università/Ente	Qualifica	Impegno	
					1° anno	2° anno
1.	DI LASCIO	Francesca Marta Lilja	Università degli Studi di BOLOGNA	Dottorando	4	4
TOTALE					4	4

7.2 - Altro personale

n°	Cognome	Nome	Università/Ente	Dipartimento	Qualifica	Impegno	
						1° anno	2° anno
1.	BEE DAGUM	ESTELA	Statistics CANADA/Universita' di Bologna	Scienze Statistiche	Dirigente/Professore ordinario a riposo	6	6
2.	GONZALEZ	DIEGO LUIS	CNR fondazione Cini	Laboratorio di acustica musicale e architettionica	Primo ricercatore	6	6
TOTALE						12	12

7.3 - Personale a contratto da destinare a questo specifico Progetto

n°	Tipologia di contratto	Costo previsto	Impegno		Note
			1° anno	2° anno	
1.	Altre tipologie	1.500	1		Analista dati ed esperto informatico per elaborazione dati e sviluppo software.
TOTALE		1.500	1	0	

7.4 - Dottorati a carico del PRIN da destinare a questo specifico Progetto

Nessuno

8 - Titolo specifico del Progetto svolto dall'Unità di Ricerca

Testo italiano

Calcolo evolutivo, ricampionamento e test di ipotesi

Testo inglese

Evolutionary computation, resampling and hypothesis testing

9 - Abstract del Progetto svolto dall'Unità di Ricerca

Testo italiano

Il progetto di ricerca dell'Unità di Bologna è incentrato sullo studio delle connessioni fra il calcolo evolutivo e i metodi di ricampionamento con particolare riferimento alla verifica di ipotesi per processi complessi. Gli ambiti applicativi su cui l'Unità di Bologna focalizzerà l'attenzione includono l'analisi delle serie storiche e la bioinformatica.

In termini qualitativi il calcolo evolutivo può essere visto come un approccio flessibile per la risoluzione di problemi motivato dai meccanismi propri della teoria dell'evoluzione [Ashlock, 2006]. Gli algoritmi genetici rientrano nell'ambito della disciplina del calcolo evolutivo. Sono strumenti che tendono a riprodurre il comportamento evolutivo di una data popolazione nella sua interazione con l'ambiente naturale in cui essa vive. Tali algoritmi sono stati pensati per risolvere problemi di ottimizzazione e il loro utilizzo in Statistica include l'ambito della selezione di modelli e della classificazione con particolare riferimento all'ambito all'apprendimento automatico e al data mining. Le connessioni fra MCMC e algoritmi genetici sono studiate in contributi recenti in letteratura [Battaglia, 2001]; tali studi aprono la strada all'utilizzo del calcolo evolutivo nell'ambito del ricampionamento con particolare riferimento ai dati dipendenti e a processi complessi.

Il ricampionamento per dati dipendenti si configura come un tema moderno e di grande rilevanza scientifica soprattutto per il suo campo di applicabilità potenzialmente senza confini. Infatti, il bootstrap classico nella sua versione parametrica o non parametrica è valido sotto l'assunzione di indipendenza, assunzione che risulta non verificata in moltissime situazioni. Per una rassegna critica si veda il numero speciale di Statistical Science dedicato al tema [Holmes, 2003].

Un possibile importante punto di contatto ai fini del progetto fra i metodi di ricampionamento e il calcolo evolutivo è costituito dal metodo dei surrogate. Tale metodo, introdotto per la prima volta agli inizi degli

anni 90 [Theiler et al., 1992] e' una tecnica Monte Carlo che permette di ottenere test di ipotesi (tipicamente di non linearita') pur in assenza di risultati sulla distribuzione delle statistiche test utilizzate. La peculiarita' del metodo consiste nel generare nuove serie temporali (dati surrogati) coerenti con l'ipotesi nulla posta e che possiedono alcune delle caratteristiche della serie originaria, quali ad esempio lo spettro di Fourier. Alcuni recenti proposte presentate in Schreiber (1998), Schreiber e Schmitz (2000) e ripreso successivamente in Giannerini Maasoumi e Bee Dagum (2007) vedono la generazione di surrogati come un problema di ottimizzazione stocastica vincolata che viene risolta per mezzo dell'algoritmo dell'annealing simulato.

Uno degli obiettivi primari dell'unita' di Bologna risiede nell'esplorare da un punto di vista teorico la possibile estensione degli algoritmi discussi per un loro utilizzo nell'ambito del ricampionamento per processi complessi con particolare riferimento alla verifica di ipotesi. Si studiera' come trasformare gli algoritmi di ottimizzazione, che sono stati pensati per la ricerca di massimi o minimi di funzioni, in algoritmi che generano da una distribuzione di probabilita' assegnata.

La validita' dei metodi proposti sara' esplorata anche in termini applicativi negli ambiti dell'analisi delle serie storiche non lineari e caotiche, la bioinformatica, l'analisi di serie storiche economiche con particolare riferimento alle matrici di smoothing e al problema dell'estrazione del segnale.

Per quanto riguarda l'analisi delle serie storiche non lineari e caotiche uno degli obiettivi del progetto e' quello di studiare l'utilizzo di misure funzionali dell'entropia per lo sviluppo di test non parametrici di non linearita' [Giannerini, Maasoumi e Bee Dagum, 2007; Granger, Maasoumi e Racine, 2004]. Inoltre si studieranno misure di sensibilita' alle condizioni iniziali per serie storiche, misure che permettono di caratterizzare la predicibilita' globale e locale di un processo non lineare [Giannerini e Rosa, 2004; Chan e Tong, 2001].

Per quanto riguarda la bioinformatica l'obiettivo del progetto dell'Unita' di Bologna prevede in primis l'applicazione dei metodi proposti all'analisi di sequenze genomiche con particolare riferimento alla struttura secondaria delle proteine. In secondo luogo, l'attenzione sara' rivolta allo studio dei GA nell'ambito della verifica di ipotesi per la classificazione delle unita' statistiche e per la selezione di modelli copula [Di Lascio e Giannerini, 2007].

Testo inglese

The research project of the University of Bologna Unit is concerned with the study of the connections between evolutionary computation and resampling techniques, with particular attention to hypothesis testing for complex processes. Regarding the applications, the Unit of Bologna will focus on time series analysis as well as Bioinformatics.

From a qualitative perspective, evolutionary computation (EC) may be viewed as a flexible approach for the solution of problems, motivated by mechanisms peculiar of the theory of evolution [Ashlock, 2006]. Genetic algorithms (GA) are part of EC. They are instruments that tend to reproduce the evolutionary behaviour of a given population in its interaction with the natural environment where it lives. These algorithms are meant to solve optimisation problems and their use in statistics includes model selection and classification with particular reference to automatic learning and data mining. Recently, the connections between Monte Carlo Markov Chains (MCMC) and GA have been investigated [Battaglia, 2001]; such contributions open the way to the use of EC within resampling methods, in particular, for dependent data and complex processes.

Resampling dependent data constitutes one of the most relevant and actual research areas in the international scientific community, mainly for its apparently infinite potential of application. In fact, the classic bootstrap, either parametric or non parametric, is valid under the hypothesis of i.i.d. data, which turns out to be untrue in a wide range of situations. For a critical review see the special number of Statistical Science [Holmes, 2003].

A possible, important, connection between EC and resampling techniques is constituted by the method of surrogate data. Such method, first introduced at the beginning of the Nineties [Theiler et al., 1992], is a Monte Carlo technique for building hypothesis testing (typically for (non) linearity) in absence of available results on the probability distribution of the test statistics involved. The peculiarity of the method consists in generating new time series (the surrogate data) that are consistent with the null hypothesis and that possess some of the characteristics of the original series, such as, for example, the Fourier spectrum. Some recent proposals presented in Schreiber (1998), Schreiber and Schmitz (2000) and then studied in Giannerini, Maasoumi and Bee Dagum (2007) introduce the generation of surrogate time series as a problem of constrained stochastic optimisation that can be solved through simulated annealing.

One of the main aims of the Unit of Bologna consists in the theoretical analysis of the possible extensions of the discussed algorithms in the context of resampling complex processes, with particular attention to hypothesis testing. One of the primary issues under investigation is how to transform optimisation algorithms, which have been thought for finding functions' maxima and minima, into algorithms that generate from a given probability distribution.

The reliability of the proposed methods will be explored in terms of applicability in the fields of non linear and chaotic time series analysis and Bioinformatics.

As concerns non linear and chaotic time series, one of the main purposes of the project consists in studying the use of functional entropy measures for the development of non parametric tests of non linearity [Giannerini, Maasoumi and Bee Dagum, 2007; Granger, Maasoumi and Racine, 2004]. Furthermore, sensitivity measures to initial conditions for time series that characterize global and local predictability of non linear processes will be studied [Giannerini and Rosa, 2004; Chan and Tong, 2001].

Regarding Bioinformatics, the first aim of the project is that of applying the methods considered above to the analysis of genomic sequences with reference to the secondary structure of proteins. Secondly, the attention will be focused on the study of GA in testing hypothesis for cluster analysis and for the selection of copula functions [Di Lascio and Giannerini, 2007].

10 - Parole chiave

n°	Parola chiave (in italiano)	Parola chiave (in inglese)
1.	CALCOLO EVOLUTIVO	EVOLUTIONARY COMPUTATION
2.	ALGORITMI GENETICI	GENETIC ALGORITHMS
3.	METODI DI RICAMPIONAMENTO	RESAMPLING TECHNIQUES
4.	BOOTSTRAP	BOOTSTRAP
5.	SERIE TEMPORALI	TIME SERIES

11 - Stato dell'arte

Testo italiano

In termini generici il calcolo evolutivo (di seguito CE) può essere visto come un approccio flessibile e adattivo per la risoluzione di problemi, adatto in particolar modo in ambito di ottimizzazione. Tale approccio prende spunto dai meccanismi propri della teoria dell'evoluzione. Dalle origini del metodo che risalgono alla fine degli anni 50 il tema ha incominciato ad avere una diffusione più ampia a partire dagli anni 80, grazie soprattutto alla disponibilità di potenza di calcolo [Ashlock, 2006].

La maggior parte delle moderne implementazioni di algoritmi evolutivi discendono da tre approcci strettamente connessi fra loro ma che sono stati sviluppati indipendentemente: gli algoritmi genetici, la programmazione evolutiva e le strategie evolutive. Per quanto inizialmente legato all'informatica, il carattere interdisciplinare delle tematiche trattate vede il calcolo evolutivo come oggetto di interesse da parte della comunità statistica a partire dagli anni 90. In particolare, si incominciano a studiare da un punto di vista statistico/probabilistico le possibili applicazioni del CE per la selezione di modelli, la classificazione, l'analisi dei gruppi e a molte delle tecniche che possono essere classificate come facenti parte di quella disciplina a cavallo fra statistica e informatica comunemente chiamata data mining. Per converso, l'adozione e lo studio del CE da parte della comunità scientifica statistica ha determinato la sistematizzazione rigorosa da un punto di vista teorico di molte delle tecniche usate [Battaglia, 2001].

Gli algoritmi genetici si configurano come strumenti che tendono a riprodurre il comportamento evolutivo di una data popolazione nella sua interazione con l'ambiente naturale in cui essa vive. Contributi recenti in letteratura [Battaglia, 2001] esplorano le connessioni fra MCMC e algoritmi genetici e suggeriscono la possibilità che il CE possa essere utilizzato non solo in ambito di ottimizzazione ma anche per la generazione di numeri casuali provenienti da una data distribuzione di probabilità aprendo la strada anche all'utilizzo in ambito di ricampionamento con particolare riferimento ai dati dipendenti e all'analisi delle serie storiche.

Il ricampionamento per dati dipendenti costituisce uno dei temi di ricerca di maggiore rilevanza e attualità nella comunità scientifica statistica internazionale. Il campo di applicabilità è enorme e risulta rilevante in tutti i problemi per cui non esiste una teoria sulla distribuzione delle quantità oggetto di inferenza o i risultati asintotici non sono applicabili. Prendendo le mosse dal bootstrap classico per dati iid nella sua versione parametrica o non parametrica gli sforzi si sono concentrati in diverse direzioni [Holmes, 2003].

Alcuni autori hanno esteso il bootstrap parametrico sui residui al caso di processi lineari di tipo ARMA [Kreiss e Franke, 1992]. Successivamente, l'interesse si è spostato verso il bootstrap non parametrico con l'idea del filtraggio tramite processi AR che ha dato luogo al Sieve bootstrap [Buhlmann, 1997] o con l'idea del Cholesky factor bootstrap [Berkovitz e Diebold, 1998] che non pone assunzioni distributive sul processo che ha generato i dati. Altri contributi rilevanti includono il moving block bootstrap [Kunsch, 1989] assieme alle sue varianti quali lo stationary bootstrap [Politis e Romano, 1994] e il blocks of blocks bootstrap [Politis, 2003]. Infine, recentemente, diversi autori hanno studiato metodi di ricampionamento che operano nel dominio delle frequenze. Tali metodi sfruttano il fatto che anche per dati dipendenti sussistono delle relazioni di indipendenza nel dominio frequenziale tali per cui è possibile ricondursi al caso iid per cui è possibile applicare i metodi classici di ricampionamento [Berkovitz e Diebold, 1998; Hidalgo, 2003].

Un ambito di ricerca strettamente legato al ricampionamento per processi complessi e di recente introduzione è costituito dal metodo dei surrogati. Tale metodo, introdotto per la prima volta agli inizi degli anni 90 [Theiler et al., 1992] è una tecnica Monte Carlo che permette di ottenere test di ipotesi (tipicamente di non linearità) pur in assenza di risultati sulla distribuzione delle statistiche test utilizzate. La peculiarità del metodo consiste nel generare nuove serie temporali (dati surrogati) coerenti con l'ipotesi nulla posta e che possiedono alcune delle caratteristiche della serie originaria, quali ad esempio lo spettro di Fourier.

Testo inglese

Evolutionary computation (EC) may be viewed as a flexible and adaptive approach for the solution of optimisation problems. Such an approach is motivated by the mechanisms that belong to evolution theory. From the very origins of the method, that date back to the end of the Fifties, the subject underwent wide diffusion in the Eighties, primarily thanks to the increased availability of computing power [Ashlock, 2006].

The majority of evolutionary algorithms modern implementations follow from three strictly connected but independently developed approaches: genetic algorithms (GA), evolutionary programming and evolutionary strategies. Although EC was initially developed within the field of Informatics, its multi disciplinary character made EC interesting and appealing to the whole scientific community. Indeed, starting from the Nineties, EC began to be studied from a probabilistic and statistical viewpoint, e.g. for model selection, classification, cluster analysis, and many other techniques that can be classified as part of data mining, a discipline that lies between Statistics and Informatics. On the other hand, the fact that the statistical scientific community started adopting and implementing EC techniques determined a rigorous and systematic theoretical study of the methods [Battaglia, 2001].

Genetic algorithms tend to reproduce the evolutionary behaviour of a given population in its interaction with the natural environment where it lives. Such a population is constituted by a given number of individuals and

the information relative to each individual is contained in his chromosomes. Typically, a chromosome is represented by a vector of units, the genes, and each gene is characterised by its position within the chromosome (locus) and by the numerical value that it assumes (allele). Adaptation of individuals to the environment is measured through a fitness function, which is in general real and positive. The population evolution from generation to generation takes place through the sequential application of genetic operators such as selection, mutation and cross over.

Recent contributions [Battaglia, 2001] explore the connections between Monte Carlo Markov chain (MCMC) and genetic algorithms and suggest the possibility that EC may be used not only in optimisation problems but also in generating random numbers from a given probability distribution, opening the way to their use within resampling methods and in particular with reference to dependent data and time series analysis.

Resampling dependent data is one of the most relevant and modern research areas in the international scientific community. The application field of applications is wide and relevant in all those problems where a distribution theory for the quantities that are object of inference is not available or where asymptotic results are not applicable. Moving from the classic bootstrap for i.i.d. data, either parametric or non parametric, the research has moved towards different directions [Holmes, 2003].

Some authors have extended the parametric residual-based bootstrap to the case of linear processes belonging to the class of ARMA models [Kreiss and Franke, 1992]. Subsequently, the interest has moved to the non parametric bootstrap associated to the idea of filtering through AR processes which led to the Sieve bootstrap [Buhlmann, 1997] or, alternatively, to the Cholesky factor bootstrap [Berkowitz and Diebold, 1998] that makes no distributional assumptions on the data generating process. Other relevant contributions include the moving block bootstrap [Kunsch, 1989], together with its variants such as the stationary bootstrap [Politis and Romano, 1994] and the blocks of blocks bootstrap [Politis, 2003]. Also, recently, several authors have concentrated on resampling techniques in the frequency domain. These methods motivate from the fact that for dependent data, some independency relations can be exploited in the frequency domain so that it is possible to apply classical resampling techniques [Berkowitz and Diebold, 1998; Hidalgo, 2003].

A recent research topic strictly linked to resampling techniques for complex processes is the method of surrogate data. Such method, first introduced at the beginning of the Nineties [Theiler et al., 1992], is a Monte Carlo technique for building hypothesis testing (typically for (non) linearity) in absence of a distributional theory of the test statistics that are employed. The peculiarity of the method consists in generating new time series (the surrogate data) that are consistent with the null hypothesis and that share with the original series some of its features, such as, for example, the Fourier spectrum.

12 Riferimenti bibliografici

- Ashlock, D., (2006). *Evolutionary computation for modeling and optimization*. New York, Springer.
- Battaglia, F., (2001). "Genetic algorithms, pseudo-random numbers generators, and Markov chain Monte Carlo methods", *Metron*, LIX (1-2), 129-154.
- Berkowitz, J., e Diebold, F.X., (1998). "Bootstrapping multivariate spectra", *Review of Economics and Statistics*.
- Buhlmann, P., (1997). "Sieve bootstrap for time series", *Bernoulli*, 3, 123-148.
- Chan, K.S., e Tong, H., (2001). *Chaos: a statistical perspective*. New York, Springer Verlag.
- Di Lascio, F.M.L., e Giannerini, S. (2007). "A new copula-based approach for clustering". Working paper.
- Fermanian, J.-P., (2005). "Goodness-of-fit tests for copulas", *Journal of Multivariate Analysis*, 95, 119-52.
- Giannerini, S., e Rosa, R., (2004). "Assessing chaos in time series: statistical aspects and perspectives", *Studies in Nonlinear Dynamics and Econometrics*, 8 (2) Article 11.
- Giannerini, S., Maasoumi, E., e Bee Dagum E., (2007). "Entropy testing for nonlinearity in time series", *Bulletin of the International Statistical Institute*, 56th session.
- Gonzalez, D.L., (2004). "Can the genetic code be mathematically described?", *Medical Science Monitor*, 10 (4), 11-17.
- Gonzalez, D.L., Giannerini, S., e Rosa, R., (2006). "A new insight into DNA: detecting structure in parity binary sequences", *IEEE Engineering in Medicine and Biology Magazine*, 25 (1), 69-81.
- Goswami, G., e Liu, J.S., (2007). "On learning strategies for evolutionary Monte Carlo", *Statistics and Computing*, 17, 23-38.
- Granger, C.W., Maasoumi, E., e Racine, J. (2004). "A dependence metric for possibly nonlinear processes", *Journal of time series analysis*, 25 (5), 649-669.
- Hidalgo, J., (2003). "An alternative bootstrap to moving blocks for time series regression models", *Journal of Econometrics*, 117, 369-399.
- Holland, J., (1995). *Hidden order, how adaptation builds complexity*, Perseus Books, Reading, Mass.
- Holmes, S., (2003). "Bootstrapping Phylogenetic Trees: Theory and Methods", *Statistical Science*, 18 (2), 241-255.
- Jasra, A., Stephens, D.A., e Holmes, C.C., (2007). "On population-based simulation for static inference", *Statistics and Computing*, 17, 263-279.
- Kreiss, J.P., e Franke, J., (1992). "Bootstrapping Stationary Autoregressive Moving-Average models", *Journal of Time Series Analysis*, 13, 287-317.
- Kunsch, H.R., (1989). "The Jackknife and the Bootstrap for General Stationary Observations", *Annals of Statistics*, 17, 1217-1241.

- Mignani, S., Giannerini, S., e Gonzalez, D. L., (2000). "Test di non linearità per la ricerca di un'eventuale dinamica caotica nel sistema cardiovascolare", *Statistica*, anno LXIV, 4, 515-522.
- Politis, D.N., e Romano, J.P., (1994). "The stationary bootstrap", *Journal of the American Statistical Association*, 89, 1303-1313.
- Politis, D.N. (2003). "The Impact of Bootstrap Methods on Time Series Analysis", *Statistical Science*, 18(2) 219-230.
- Schreiber, T., e Schmitz, A., (1996). "Improved surrogate data for nonlinearity tests", *Physical Review Letters*, 77(4), 635-638.
- Schreiber, T., (1998). "Constrained randomization of time series data", *Physical Review Letters*, 90(10), 2105-2108.
- Schreiber, T., e Schmitz, A., (2000). "Surrogate time series", *Physica D*, 142 (3-4), 346-382.
- Small, M., e Judd, K., (1998). "Correlation dimension: A pivotal statistic for non-constrained realizations of composite hypotheses in surrogate data analysis", *Physica D*, 120 (3-4), 386-400.
- Small, M., Judd, K., e Mees, A., (2001). "Testing time series for nonlinearity", *Statistics and Computing*, 11, 257-268.
- Small, M., (2005). "Applied nonlinear time series analysis", World Scientific, Singapore. Applications in Physics, Physiology and Finance.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., e Farmer, J.D., (1992). "Testing for nonlinearity in time series: the method of surrogate data", *Physica D*, 58, 77-94.

13 - Descrizione del programma e dei compiti dell'Unità di Ricerca

Testo italiano

Il presente progetto di ricerca si propone di studiare le connessioni che sussistono fra il calcolo evolutivo (di seguito CE), con particolare riferimento agli algoritmi genetici (di seguito GA), e i metodi di ricampionamento. Come menzionato in precedenza il calcolo evolutivo in ambito statistico viene primariamente utilizzato nell'ambito dell'analisi dei gruppi, della selezione di modelli e della classificazione. Recentemente diversi autori hanno incominciato a esplorarne la sua applicabilità in ambito MCMC dando luogo a quello che viene comunemente chiamato MCMC evolutivo; si veda ad esempio Battaglia (2001) e Goswami e Liu, (2007). Inoltre, l'introduzione dei concetti propri del CE e il loro utilizzo al fine di migliorare le prestazioni degli algoritmi MCMC viene discussa in Jasra et al. (2007). Si parla ad esempio di "population based simulation" ove invece di lavorare su singole realizzazioni di catene di Markov si considera l'evoluzione di più catene in parallelo; o ancora si usano i GA per migliorare le proprietà di mixing delle catene.

Questi contributi recenti suggeriscono la possibilità che il CE possa essere utilizzato non solo nel campo dell'ottimizzazione o negli ambiti sopra citati ma anche per la generazione di realizzazioni da una data distribuzione di probabilità e aprono la strada anche al suo impiego nell'ambito dei metodi di ricampionamento per fenomeni complessi. Una chiara segnale in questa direzione viene dall'analisi delle serie storiche non lineari, motivata dalla teoria del caos, con il metodo dei surrogati. Il metodo dei surrogati, introdotto in Theiler et al. (1992) può essere visto come un approccio basato sul ricampionamento per costruire una verifica di ipotesi in assenza della teoria riguardante la distribuzione delle statistiche di interesse. Il tratto caratteristico dell'approccio consiste nel generare serie storiche, dette serie surrogate, che siano coerenti con una data ipotesi nulla. Il secondo passo consiste nella scelta di una statistica test che abbia potere discriminatorio nei confronti dell'ipotesi postulata. Si menzionano, ad esempio, l'errore di previsione non lineare, misure basate sull'entropia, gli esponenti di Lyapunov e misure basate sull'integrale di correlazione. Infine, la significatività del test può essere valutata in modo standard confrontando il valore della statistica calcolata sulla serie originale con i valori ottenuti dai surrogati.

Nella sua prima formulazione il metodo è stato introdotto per testare l'ipotesi di non linearità della serie per cui l'ipotesi nulla è che il processo che ha generato i dati sia un processo lineare. In questo caso sono stati generati surrogati che possiedono lo stesso spettro dei dati originali attraverso la randomizzazione della fase della trasformata di Fourier dei dati seguita da una procedura di aggiustamento. Studi successivi hanno mostrato come un tale approccio risulti poco efficace in quanto porta a serie distorsioni in termini di potenza e ampiezza dei test [Schreiber e Schmitz, 1996 e 2000].

Estensioni del metodo in varie direzioni sono state proposte in seguito. In generale, tali estensioni coinvolgono l'introduzione di nuove statistiche test o di algoritmi ad hoc per generare surrogati coerenti con una data ipotesi, non necessariamente di linearità. Ad esempio in Small e Judd (1998), Small et al. (2001), Small (2005) si discute il problema delle statistiche pivotali e si introduce una classe di statistiche basate sull'integrale di correlazione. È importante notare che nonostante il metodo sia stato studiato in diverse situazioni sembrano mancare studi sistematici sulle prestazioni dei test così ottenuti.

Il punto di contatto fra il metodo dei surrogati, il calcolo evolutivo e i metodi di ricampionamento è costituito dall'approccio presentato in Schreiber (1998), Schreiber e Schmitz (2000) e ripreso successivamente in Giannerini, Maasoumi e Bee Dagum (2007). In breve, la generazione di surrogati è vista come un problema di ottimizzazione stocastica vincolata che viene risolta per mezzo dell'algoritmo dell'annealing simulato. Le caratteristiche che i surrogati devono possedere sono espresse in termini di vincoli i quali al loro volta definiscono una funzione di costo che assume il suo valore minimo globale quando i vincoli sono soddisfatti. Successivamente, si minimizza la funzione di costo nello spazio delle permutazioni dei dati osservati per mezzo dell'annealing ottenendo così una serie surrogate che appare come una permutazione casuale della serie originale avente in comune con quest'ultima le caratteristiche definite per mezzo della funzione di costo.

Diversi studi hanno messo in luce la superiorità dello schema proposto rispetto ai contributi precedenti. In Mignani, Giannerini e Gonzalez (2000) l'approccio è utilizzato per la verifica del carattere non lineare di serie storiche relative alla dinamica cardiaca mentre in Giannerini, Maasoumi e Bee Dagum (2007) si studiano gli aspetti teorici e pratici di un test di non linearità per serie storiche che combina il metodo con un

statistica test basata su di una misura legata all'entropia.

La caratteristica peculiare di un tale approccio risiede nel fornire uno schema unificato di generazione dei surrogate che risulti indipendente dall'ipotesi da testarsi. Infatti, per cambiare l'ipotesi in esame e' sufficiente utilizzare diverse funzioni di costo. Inoltre, la scelta dell'annealing simulato non e' l'unica possibile. Le considerazioni fatte aprono le porte allo studio dell'utilizzo degli algoritmi di ottimizzazione stocastica e in generale di tutti gli algoritmi che fanno capo al calcolo evolutivo per la generazione di dati con caratteristiche assegnate.

Uno degli obiettivi primari dell'unita' di Bologna risiede nell'esplorare sia dal punto di vista teorico sia in termini di applicabilita' la possibile estensione degli algoritmi discussi per un loro utilizzo nell'ambito del ricampionamento per processi complessi, non strettamente legato, quindi, alla verifica di ipotesi per serie temporali. In particolare, si studiera' come trasformare gli algoritmi di ottimizzazione, che sono stati pensati per la ricerca di massimi o minimi di funzioni, in algoritmi che generano da una distribuzione di probabilita' assegnata. E' importante notare che mentre nel caso dell'MCMC tale distribuzione e' nota nella forma nel caso del ricampionamento la distribuzione che ha generato i dati osservati (di seguito DGP) e' ignota e' nota che assieme al problema dell'effettiva appartenenza dei dati ricampionati alla distribuzione voluta costituiscono i punti cardine delle tecniche di ricampionamento.

Una soluzione che verra' presa in considerazione e' quella della generazione di dati ricampionati che condividano con i dati sperimentali funzionali legati al DGP. In analogia con il metodo dei surrogate che utilizza l'annealing simulato e incorporando le idee provenienti dal CE verranno esplorati i seguenti temi: i) utilizzare funzioni di costo/fitness multidimensionali e studiare l'impatto dell'introduzione contemporanea di piu' vincoli; ii) prendendo spunto dalle motivazioni sottostanti i primi tentativi di formalizzare i GA [Holland, 1995] ove l'oggetto di studio e' l'evoluzione di un sistema adattivo complesso si studiera' l'utilizzo di funzioni di costo/fitness adattive, cioe' che cambiano al passare del tempo (o delle generazioni). I legami con il campionamento ad importanza utilizzato nel MCMC sono chiari; iii) Dal momento che in ambito di ricampionamento si cerca di riprodurre la distribuzione di una statistica di interesse si studiera' se e' possibile utilizzare a tal fine metodi di simulazione di tipo "population based" in cui si studia l'evoluzione parallela di tutti gli individui e si valtera' se un tale approccio e' preferibile rispetto ad altre soluzioni.

La validita' dei metodi proposti sara' saggiata sia in termini teorici sia in termini applicativi. Per quanto riguarda quest'ultimo aspetto gli ambiti di applicazione saranno l'analisi delle serie storiche non lineari e caotiche, la bioinformatica, l'analisi di serie storiche economiche con particolare riferimento alle matrici di smoothing e al problema dell'estrazione del segnale.

Per quanto attiene al tema dell'analisi delle serie storiche non lineari e caotiche uno degli obiettivi del progetto e' quello di studiare l'utilizzo di misure funzionali dell'entropia sviluppati nel contesto della teoria dell'informazione per lo sviluppo di test non parametrici di non linearita' che non siano legati ad assunzioni sul processo generatore dei dati, che abbiano un grande spettro di applicabilita' e, allo stesso tempo, possiedano buone proprieta' di potenza e ampiezza [Granger, Maasoumi e Racine, 2004; Giannerini, Maasoumi e Bee Dagum, 2007].

Un ulteriore aspetto che sara' oggetto di studio riguarda le misure di sensibilita' alle condizioni iniziali per serie storiche. Tali misure costituiscono un'estensione della nozione di sensibilita' alle condizioni iniziali motivata dalla teoria del caos e ben definita per processi deterministici. Lo studio di tali misure in termini di momenti condizionali e' stata effettuata in Chan e Tong (2001) e in Giannerini e Rosa (2004). Uno degli obiettivi del progetto e' l'estensione di tali misure a tutta la distribuzione condizionata. E' importante ricordare che questi indicatori, in quanto misure dell'interazione complessa fra segnale e rumore, caratterizzano la predicibilita' locale di un processo non lineare, non necessariamente legata alla quantita' di rumore presente nel sistema.

Infine, il progetto prevede l'implementazione di una libreria software originale per le serie storiche non lineari che raccolga e renda disponibile alla comunita' scientifica internazionale la maggior parte degli algoritmi sviluppati all'interno del progetto. La libreria chiamata tseriesChaos e' parte della suite R e una sua versione preliminare e' gia' disponibile sul sito del progetto R.

Per quanto riguarda la bioinformatica l'obiettivo del progetto dell'Unita' di Bologna e' duplice. In primis ci si propone di applicare i metodi proposti all'analisi di sequenze genomiche muovendo da uno studio recente sulla struttura matematica del codice genetico [Gonzalez, 2004]. Alcuni risultati preliminari [Gonzalez, Giannerini e Rosa, 2006] sembrano supportare l'ipotesi dell'esistenza di un meccanismo di correzione dell'errore all'interno del codice genetico. La rappresentazione in stringhe binarie degli aminoacidi determinata dalla codifica di Gonzalez (2004) apre la porta all'utilizzo dei GA in questo ambito e la palindromia emergente da tale rappresentazione induce interrogativi relativi al suo ruolo nella codifica numerica delle sequenze proteiche. L'attenzione verra' focalizzata sulle proteine con l'intento di caratterizzare la loro struttura secondaria a partire dal modello matematico proposto.

In secondo luogo, l'attenzione sara' rivolta allo studio dei GA nell'ambito della verifica di ipotesi per la classificazione delle unita' statistiche e per la selezione di modelli copula. Il punto di partenza e' costituito da un nuovo algoritmo di clustering basato sulle funzioni copula [Di Lascio e Giannerini, 2007] che consiste in una procedura iterativa capace di individuare il corretto numero di clusters e di classificare le osservazioni rispettando la struttura di dipendenza del DGP. L'algoritmo proposto si basa, infatti, su un modello per la funzione copula che deve essere scelto dall'utente. La definizione di criteri per la selezione del migliore modello copula costituisce oggi una fervida area di ricerca [Ferriani, 2005] in cui, secondo la nostra conoscenza, non sono stati mai utilizzati i GA.

Testo inglese

The main aim of this research project is to study the connections between evolutionary computation (EC), with particular reference to genetic algorithms (GA), and resampling techniques. As it has been mentioned above, EC in statistics is primarily used in the context of cluster analysis, model selection and classification. Recently, several authors started to explore its applicability within the theory of MCMC, leading to what is commonly known as evolutionary MCMC; see for example Battaglia (2001) and Goswami and Liu (2007). The concepts of EC and their use to improve the performances of MCMC algorithms are discussed in Jasra et al. (2007). For example, "population based simulations" are considered wherever instead of working on single realisations of Markov chains, the evolution of a set of parallel chains is considered; also, GA may be employed to improve the mixing properties of Markov chains.

All the recent contributions on EC, suggest that the latter may be used not only for optimisation or in relation to MCMC but also for generating realisations from a given probability distribution, opening the way to their employment in resampling methods for complex phenomena. A clear indication in this direction comes

from the analysis of non linear time series with the method of surrogates, motivated by chaos theory. The surrogate method, introduced in Theiler et al. (1992) can be seen as a resampling-based approach to hypothesis testing in absence of a theory on the distribution of the test statistics of interest. The peculiar trait of this approach consists in, first, generating surrogate time series that are consistent with a given null hypothesis. The second step consists in choosing a test statistics that has discriminating power with respect to the null hypothesis. For example, non linear forecast error, entropy-based measures, Lyapunov exponents and correlation integral-based measures. Finally, the significance of the test may be evaluated through standard procedure by comparing the value of the test statistics computed on the original series with the values obtained from the surrogates.

In its original formulation, the method has been introduced to test the hypothesis that the series is not linear, i.e. the null hypothesis is that of a linear data generating process. In this setting, surrogates that possess the same spectrum of the original data have been generated through randomisation of the phase of the Fourier transform, followed by an adjustment procedure. Further studies have shown that such an approach is severely biased in terms of power and size of the test [Schreiber and Schmitz, 1996 and 2000].

Many generalisations of the surrogate method in different directions have been proposed. In general, such extensions introduce new test statistics or ad hoc algorithms to generate surrogates that are consistent with a given hypothesis, not necessarily of non linearity. In Small e Judd (1998), Small et al. (2001), Small (2005) the problem of pivotal statistics is discussed and a class of statistics based on the correlation integral is proposed. It is important to remark that even if the method has been applied in many different situations, no systematic study is available on the test performances.

A clear link among surrogate methods, EC and resampling techniques is introduced in Schreiber (1998), Schreiber and Schmitz (2000) and in Giannerini, Maasoumi and Bee Dagum (2007). Generation of surrogates is seen as constrained stochastic optimisation problem that is solved through the simulated annealing algorithm. The surrogate series must possess characteristics that are expressed as constraints that define a loss function that has its global minimum when such constraints are fulfilled. Successively, the cost function is minimised over the space of the permutations of the observed data through simulated annealing, giving birth to a surrogate series, namely, a random permutation of the original series that share the characteristics defined by the loss function.

Several studies evidenced the superiority of this scheme with respect to the previous contributions. In Mignani, Giannerini and Gonzalez (2000) the approach has been used to test non linearity in the cardiovascular dynamics, whereas Giannerini, Maasoumi and Bee Dagum (2007) discuss the theoretical and empirical aspects of a non linearity test for time series that combines the surrogate method with a test statistic based on an entropy measure.

The most important aspect of such an approach is the fact that it provides a unified method for generating surrogates independent from the hypothesis to be tested. In fact, in order to change the null hypothesis, it is sufficient to modify the loss function. Furthermore, simulated annealing is not the only admissible choice. These considerations open the way to the analysis of stochastic optimisation algorithms and in general to EC algorithms for generating data with given characteristics.

Indeed, one of the main objectives of the Unit of Bologna consists in the theoretical as well as empirical analysis of the generalisation of these algorithms in relation to resampling complex processes with particular emphasis on hypothesis testing. In particular, it will be studied the possibility of treating optimisation algorithms into algorithms that generate a from a given probability distribution. It is important to remark that, while in MCMC the functional form of the distribution is known, in the case of resampling problems the data generating process/distribution (DGP) is unknown. This matter, together with the study whether resampled data effectively belong to a given distribution constitute the fundamental issues of resampling methods.

One solution that will be considered is that of generating resampled data that share, with the original data, functionals connected to the DGP. In analogy with the surrogate method that uses the simulated annealing and together with the ideas of EC, the following problems will be deeply analyzed:

i) the use of multidimensional cost/fitness functions and the simultaneous study of a set of constraints; ii) the use of adaptive cost/fitness functions, that change with time (or generations of populations), motivated by the initial formalisation of GA [Holland, 1995] where the object of interest is the evolution of a complex system. The relationship with importance sampling methods in MCMC are clear; iii) motivated by the fact that the aim of resampling techniques is to reproduce the distribution of a statistics of interest we will study "population based" simulation methods where the parallel evolution of all the individuals is considered.

The reliability of the proposed methods will be evaluated both from a theoretical and an applied perspective. With regards to this latter aspect, the application areas will be the analysis of non linear and chaotic time series as well as Bioinformatics.

As concerns non linear and chaotic time series, one of the main purposes of the project consists in studying the use of functional entropy measures for the development of non parametric tests of non linearity that do not depend on assumptions to the data generating process, have a wide range of applications and that possess good power and size properties [Granger, Maasoumi and Racine, 2004; Giannerini, Maasoumi and Bee Dagum, 2007].

A further object of study concerns sensitivity measures to initial conditions for time series. These measures extend the notion of sensitivity to initial conditions motivated by chaos theory and well defined for deterministic processes. Such measures have been explored in terms of conditional moments in Chan and Tong (2001), Giannerini and Rosa (2004). In the project we are interested in extending such measures to the whole conditional distribution. It is important to remark that these measures, as indicators of the complex interaction between signal and noise, characterise the local predictability of a non linear process which does not necessarily depends on the amount of noise present in the system.

Furthermore, the project will include the implementation of an original software library for non linear time series that collects the algorithms developed within the project and makes them available to the international scientific community. The *tseriesChaos* library is integral part of the suite *R* and an preliminary version of it is already available on the web site of the *R* project.

Concerning Bioinformatics, the first aim of the project is that of applying the methods considered above to the analysis of genomic sequences, based on a recent study on the mathematical structure of the genetic code [Gonzalez, 2004]. Some preliminary results by Gonzalez, Giannerini and Rosa (2006) support the hypothesis of existence of a error correction mechanism inside the genetic code. The representation of amino acids in

binary sequences determined by Gonzalez (2004) opens the door to the use of GA in this setting. Furthermore, the palindromic structure emerging from this representation poses questions on its role on the numeric codification of protein sequences. Attention will be focused on proteins with the aim of characterising their secondary structure starting from the proposed mathematical model.

Secondly, the attention will focus on the study of GA in testing hypothesis in classification problems and for the selection of copula functions. The starting point is a new clustering algorithm based on copula functions [Di Lascio and Giannerini, 2007] that consists in an iterative procedure capable to detect the correct number of clusters and to classify the observations taking into account their dependence structure. The proposed algorithm is based on a model for copula function that has to be chosen by the researcher. The problem of model selection for copula models is modern and important research area [Fermanian, 2005] in which, to the best of our knowledge, the use of GA have not been explored.

14 - Descrizione delle attrezzature già disponibili ed utilizzabili per la ricerca proposta

Testo italiano

n°	anno di acquisizione	Descrizione
1.	2006	<p>SERVER IBM PER IL CALCOLO SCIENTIFICO. 4 AMD Opteron dual core della serie 8218 2.6 Ghz, 8 core totali</p> <p>32 Gb di RAM ECC</p> <p>2 dischi SCSI U320 160 Gb in RAID 0</p> <p>OS: LINUX OPEN SUSE 10.2 64 bit</p> <p>Software installato</p> <p>R compilato con BLAS ACML ottimizzata per processori AMD e multithreaded. gcc 4.1.2 gfortran 4.3.0</p>

Testo inglese

n°	anno di acquisizione	Descrizione
1.	2006	<p>Server IBM for Scientific Computing</p> <p>4 AMD Opteron dual core 8218 2.6 Ghz, 8 core overall.</p> <p>32 Gb RAM ECC</p> <p>2 HD SCSI U320 160 Gb, RAID 0</p> <p>OS: LINUX OPEN SUSE 10.2 64 bit</p> <p>Software:</p> <p>R compiled and linked to multithreaded ACML BLAS. gcc 4.1.2 gfortran 4.3.0</p>

15 - Descrizione delle Grandi attrezzature da acquisire (GA)

Testo italiano

Nessuna

Testo inglese

Nessuna

16 - Mesi persona complessivi dedicati al Progetto

		Numero	Impegno 1° anno	Impegno 2° anno	Totale mesi persona
Componenti della sede dell'Unità di Ricerca		3	18	18	36
Componenti di altre Università/Enti vigilati		0			
Titolari di assegni di ricerca		0			
Titolari di borse	Dottorato	1	4	4	8
	Post-dottorato	0			
	Scuola di Specializzazione	0			

<i>Personale a contratto</i>	<i>Assegnisti</i>	0			
	<i>Borsisti</i>	0			
	<i>Altre tipologie</i>	1	1	0	1
<i>Dottorati a carico del PRIN da destinare a questo specifico progetto</i>		0	0	0	0
<i>Altro personale</i>		2	12	12	24
TOTALE		7	35	34	69

17 - Costo complessivo del Progetto dell'Unità articolato per voci

Voce di spesa	Spesa in Euro	Descrizione dettagliata (in italiano)	Descrizione dettagliata (in inglese)
Materiale inventariabile	6.000	<i>Computer, software, libri e riviste.</i>	<i>Computer, software, books and journals.</i>
Grandi Attrezzature	0		
Materiale di consumo e funzionamento	5.000	<i>2300 euro: quota forfettaria, pari al 10% del costo complessivo, per la gestione amministrativa del fondo; il restante importo per fotocopie, supporti di memorizzazione, cancelleria, spese di manutenzione, etc.</i>	<i>2300 euro: equal to the 10% of the total amount for the administrative management; the remainder is for stationery.</i>
Spese per calcolo ed elaborazione dati			
Personale a contratto	1.500	<i>Contratto per analisi di dati e/o sviluppo software</i>	<i>Data analyst and/or software development</i>
Dottorati a carico del PRIN da destinare a questo specifico progetto	0		
Servizi esterni			
Missioni	7.500	<i>Partecipazioni a incontri tra le unità costituenti il progetto, collaborazioni con ricercatori stranieri, organizzazione di seminari internazionali e partecipazione a convegni internazionali.</i>	<i>Meetings among research units, collaborations with other researchers, organisation of seminars and conferences.</i>
Pubblicazioni			
Partecipazione / Organizzazione convegni	3.000	<i>Quote di partecipazione a convegni nazionali ed internazionali, organizzazione del convegno finale relativo ai risultati del progetto.</i>	<i>Fees for attending conferences, organisation of a final workshop where the results of the research units will be presented.</i>
Altro			
TOTALE	23.000		

18 - Prospetto finanziario dell'Unità di Ricerca

Voce di spesa	Importo in Euro
a.1) finanziamenti diretti, disponibili da parte di Università/Enti vigilati di appartenenza dei ricercatori dell'unità operativa	
a.2) finanziamenti diretti acquisibili con certezza da parte di Università/Enti vigilati di appartenenza dei ricercatori dell'unità operativa	6.900
b.1) finanziamenti diretti disponibili messi a disposizione da parte di soggetti esterni	
b.2) finanziamenti diretti acquisibili con certezza, messi a disposizione da parte di soggetti esterni	
c) cofinanziamento richiesto al MUR	16.100
Totale	23.000

19 - Certifico la dichiarata disponibilità e l'utilizzabilità dei finanziamenti a.1) a.2) b.1) b.2)

SI

Firma _____

(per la copia da depositare presso l'Ateneo e per l'assenso alla diffusione via Internet delle informazioni riguardanti i programmi finanziati e la loro elaborazione necessaria alle valutazioni; D. Lgs, 196 del 30.6.2003 sulla "Tutela dei dati personali")

Firma _____

Data 23/10/2007 ore 19:13

ALLEGATO

Curricula scientifici dei componenti il gruppo di ricerca

Testo italiano

1. DI LASCIO Francesca Marta Lilja

Curriculum:

POSIZIONE ATTUALE

DOTTORANDA in "Metodologia Statistica per la Ricerca Scientifica" (XX Ciclo, Iscritta al terzo anno) presso il Dipartimento di Scienze Statistiche "P. Fortunati" dell'Università degli Studi di Bologna, sede di Bologna. Tutor: Prof.ssa Paola Monari.

TUTOR degli insegnamenti di "Analisi di serie storiche e multidimensionali" nell'ambito del Corso di Laurea Triennale in Statistica e Ricerca Sociale (Docente: Prof.ssa M. Pillati) e di "Analisi multivariata" nell'ambito del Corso di Laurea Specialistica in Scienze Statistiche (Docente: Prof.ssa A. Montanari), presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Bologna, sede di Bologna.

TUTOR degli insegnamenti di "Istituzioni di statistica" nell'ambito del Corso di Laurea Triennale in Economia e Diritto (CLED), (Docente: Prof. A. Roverato) e di "Statistica per le applicazioni giuridico-economiche" nell'ambito del Corso di Laurea Specialistica in Economia e Diritto (CLASED), (Docente: Prof. A. Roverato), presso la Facoltà di Economia dell'Università degli Studi di Bologna, sede di Bologna.

PERCORSO DI STUDI

Nel giugno 1998 ha conseguito il *DIPLOMA DI MATURITA' SCIENTIFICA* presso il Liceo Scientifico A. Gatto' (Agropoli, SA) con votazione 60/60.

Dal settembre 1999 al marzo 2004 è stata studente del Corso di Laurea in "Scienze Statistiche Demografiche e Sociali" presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Bologna, sede di Bologna.

-Superati 23 esami con un punteggio medio pari a 29.5/30 e con 10 laudi.

Nel marzo 2004 ha conseguito la *LAUREA* in "Scienze statistiche demografiche e sociali" presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Bologna, sede di Bologna, con votazione 110/110 *CUM LAUDE*.

-Titolo della tesi: *Un nuovo approccio alla caratterizzazione dei processi stocastici lineari e non lineari* (relatore: Prof.ssa Estela Bee Dagum, Università di Bologna; correlatore: Dott.ssa Alessandra Luati, Università di Bologna).

Nel gennaio 2005 è stata ammessa al dottorato in "Metodologia Statistica per la Ricerca Scientifica" presso il Dipartimento di Scienze Statistiche "P. Fortunati" dell'Università degli Studi di Bologna, sede di Bologna.

CURRICULUM SCIENTIFICO

Dal 10 al 14 maggio 2004 ha partecipato alla *SCUOLA SIS "Analisi delle serie storiche non lineari: teoria ed applicazioni"* presso il Campus Universitario di Fisciano, Salerno, Italia. Organizzata dalla SIS (Società Italiana di Statistica). Coordinatore: Prof.ssa Alessandra De Rose.

Nel dicembre 2004 ha vinto il concorso pubblico di Dottorato di Ricerca in "Metodologia Statistica per la Ricerca Scientifica" bandito dall'Università degli Studi di Bologna (Gazzetta Ufficiale n. 77 IV serie speciale, del 28.09.2004). Prima classificata con il punteggio di 114/120.

Dal 3 al 23 luglio 2005 ha partecipato al *CORSO ESTIVO DI STATISTICA E PROBABILITA' "Sequential Design and Analysis with Application to Clinical Trials"*, Torgnon (AO). Organizzato dall'IMQ (Istituto di Metodi Quantitativi) dell'Università Bocconi di Milano, Italia. Coordinatore: Prof. Donato Michele Cifarelli.

Dal 9 al 14 dicembre 2005 ha partecipato alla *5a SCUOLA INVERNALE INTERNAZIONALE* su "REASONing under PARTial Knowledge", Foligno (Perugia). Organizzata dal Dipartimento di Matematica ed Informatica dell'Università di Perugia, Italia. Coordinatore: Prof. Romano Scozzafava.

Dal 13 al 17 febbraio 2006 ha partecipato alla *7a SCUOLA INVERNALE INTERNAZIONALE* su "APPLIED BIOINFORMATICS: the Test Case of Human Genome", Bologna, Italia. Organizzata dall'AIRBBC (Associazione Italiana per la Ricerca in Biofisica e Biologia computazionale) c/o il Dipartimento di Biologia dell'Università degli Studi di Bologna, Italia. Coordinatore: Prof.ssa Rita Casadio.

Dal settembre 2006 al febbraio 2007 ha svolto un periodo di studio e ricerca come *VISITING STUDENT* presso il Dipartimento di Biologia dell'Università della Pennsylvania, Philadelphia, PA, USA. Prof. Warren J. Ewens. Titolo del progetto: "Analisi statistica delle sequenze del DNA: struttura della dipendenza e analisi dei clusters negli esperimenti microarray".

ATTIVITÀ DIDATTICA

Durante l'anno accademico 2006/2007 ha svolto *ATTIVITÀ DI SUPPORTO ALLA DIDATTICA E TUTORATO* per l'insegnamento di "Analisi di serie storiche e multidimensionali" nell'ambito del Corso di Laurea Triennale in "Statistica e Ricerca Sociale" presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Bologna, sede di Bologna. Docente: Prof.ssa M. Pillati.

Durante l'anno accademico 2006/2007 ha svolto *ATTIVITÀ DI SUPPORTO ALLA DIDATTICA E TUTORATO* per l'insegnamento di "Analisi multivariata" nell'ambito del Corso di Laurea Specialistica in "Scienze Statistiche" presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Bologna, sede di Bologna. Docente: Prof.ssa A. Montanari.

PARTECIPAZIONE A SOCIETÀ SCIENTIFICHE

Dal marzo 2004 è *Membro junior* della Società Italiana di Statistica (SIS).

Dall'ottobre 2006 è *Membro Student* dell'American Statistical Association (ASA). Sezioni di Membership: Biometrics (BIOM), Joint Stat Comp & Stat Graph (CPGH), Bayesian Statistical Science (SBSS), Teaching Stats in the Health Sciences (TSHS).

PRESENTAZIONE DI SEMINARI

Nell'ottobre del 2005 (durante il primo anno di Dottorato) ha tenuto un seminario dal titolo "Un nuovo approccio alla caratterizzazione dei processi stocastici lineari e non lineari" nell'ambito del corso di dottorato in "Processi Stocastici" del Prof. R. Rosa presso il Dipartimento di Scienze Statistiche "P. Fortunati" dell'Università degli Studi di Bologna, sede di Bologna.

CONVEGNI ORGANIZZATI

Ha fatto parte della Commissione Esecutiva del Workshop Finale del PRIN/COFIN 2002 su "Statistical Inference on the Stochastic and Deterministic Dynamics of Observed Time Series", Bressanone, Italia, 9-11 Giugno, 2005.

Pubblicazioni:

- ◆ DI LASCIO F., EWENS WJ. (2007). Comparison between K-Means and Hierarchical Clustering of Dependent and Independent Data Generated from Multivariate Gaussian Copula Function. Annual Conference of the New Zealand Statistical Association (NZSA). 4-6 Luglio. : (NEW ZEALAND).
- ◆ DI LASCIO F. (2006). A New Starting Point to Analyze the Nonlinear Time Series: the Nonlinear ARMA Doubly Stochastic and State Dependent Models. Proceedings della 5a Conferenza Internazionale Annuale su Statistics, Mathematics and Related Field. 16 - 18 Gennaio. (pp. 453-470). : (UNITED STATES).
- ◆ DI LASCIO F. (2006). Complications of Percutaneous Laser Ablation for Hepatocellular Carcinoma: A Multicenter Study in 520 Patients. 18th Congress of Ultrasound in conjunction with the XVIII Congresso Nazionale SIUMB. 15-19 Settembre. (pp. 30). : (ITALY).
- ◆ ACCOGLI, E, TITTOTO, P, ROMANICO, A, DI LASCIO F., ET AL. (2005). Accuratezza diagnostica dell'ecografia con mezzo di contrasto (CEUS) vs l'ecografia basale (US) nella caratterizzazione dell'iperplasia nodulare focale (INF). XVII Congresso nazionale della Società Italiana di Ultrasonologia in Medicina e Biologia (SIUMB). 12 -15 Novembre. (pp. 4-5). ROMA: (ITALY).
- ◆ DI LASCIO F. (2005). Nonlinear ARMA doubly stochastic and state dependent models: a new general approach to nonlinear time series analysis. 30a Conferenza su Stochastic Processes and their Applications. 26 Giugno - 1 Luglio. (pp. 32). : (UNITED STATES).
- ◆ ROSENBERGER WF, HU F, DI LASCIO F., ET AL. (2006). Randomization for Balancing Treatment Assignments in "Sequential Design and Analysis with Application to Clinical Trials". Quaderno dell'Istituto di Metodi Quantitativi (IMQ), Università Bocconi, Milano, Italia. (vol. n. 13, parte I, cap.2., pp. 13-25).

2. GUIDOTTI Laura

Curriculum:

Pubblicazioni:

- ◆ E. BEE DAGUM, GUIDOTTI L., A. LUATI. (2005). Some Statistical Applications of Centrosymmetric Matrices. (pp. 97 - 104). [RELAZIONE]. HEIDELBERG: Springer-Verlag (GERMANY).
- ◆ GUIDOTTI L. (2005). Wilde in Arcadia. (pp. 379 - 385). ISBN/ISSN: 88 470 0346 6. [RELAZIONE] La relazione, il volume, la rassegna teatrale si configurano come attività svolte nell'ambito del progetto europeo: "Diffusion and i... MILANO: Springer-Verlag (ITALY).

3. LUATI Alessandra

Curriculum:

Dal 2001, ricercatore presso la Facoltà di Scienze Statistiche dell'Università di Bologna. Titolare dei corsi triennale e specialistico di Analisi delle Serie Storiche presso la Facoltà di Scienze Statistiche.

Membro del comitato di redazione di Statistica, referee per varie riviste nazionali ed internazionali, reviewer per Mathematical Reviews.

Membro della Società Italiana di Statistica, della Società Americana di Statistica, dell'International Institute of Forecaster, del Gruppo di lavoro per l'Analisi delle serie temporali della Società Italiana di Statistica.

Ha collaborato attivamente all'Unità di Bologna coordinata dalla prof. Estela Bee Dagum (anche coordinatore nazionale) nei Progetti di ricerca finanziati dal Miur negli anni 2000, "Linearità e non linearità nelle dinamiche di serie storiche", 2002, "Inferenza statistica sulla dinamica stocastica e deterministica di serie storiche osservate" e 2004, "Stima parametrica e non parametrica e previsione delle dinamiche di momenti condizionali di serie storiche".

Gli interessi scientifici riguardano l'analisi delle serie storiche, con particolare attenzione ai metodi di stima non parametrica del segnale, e l'inferenza statistica per i sistemi dinamici quantistici, con particolare riferimento alla teoria dell'informazione di Fisher in detti sistemi.

Pubblicazioni:

- ◆ MANFRINI O, BAZZOCCHI G, LUATI A., BORGHI A, MONARI P, BUGIARDINI R. (2006). Coronary spasm reflects inputs from the adjacent esophageal system. AMERICAN JOURNAL OF PHYSIOLOGY. vol. 290, pp. H2085 - H2091 ISSN: 0363-6135. [ARTICOLO].
- ◆ LUATI A., G. TASSINARI. (2005). Intervention Analysis to Identify Significant Exposures in Pulsing Advertising Campaigns: an Operative Procedure. COMPUTATIONAL MANAGEMENT SCIENCE. vol. 4, n. 4, pp. 295 - 308 ISSN: 1619-697X. [ARTICOLO].
- ◆ E. BEE DAGUM, LUATI A. (2004). Relationship between local and global nonparametric estimators measures of fitting and smoothing. STUDIES IN NONLINEAR DYNAMICS AND ECONOMETRICS. vol. vol. 8 n. 2 art. 17 ISSN: 1081-1826. [ARTICOLO].
- ◆ E.B. DAGUM, LUATI A. (2004). A linear transformation and its properties with special applications in time series filtering. LINEAR ALGEBRA AND ITS APPLICATIONS. vol. 338, pp. 107 - 118 ISSN: 0024-3795. [ARTICOLO].
- ◆ LUATI A. (2004). Maximum Fisher Information in Mixed State Quantum Systems. ANNALS OF STATISTICS. vol. 32, 4, pp. 1770 - 1779 ISSN: 0090-5364. [ARTICOLO].
- ◆ LUATI A. (2003). A real formula for transition probabilities. STATISTICA. vol. 1, pp. 71-77 ISSN: 0390-590X.
- ◆ E.B. DAGUM, LUATI A. (2002). Global and local statistical properties of fixed-length nonparametric smoothers. STATISTICAL METHODS & APPLICATIONS. vol. 11, 3, pp. 313-333 ISSN: 1618-2510.
- ◆ LUATI A., P. PARUOLO. (2002). Sulla distribuzione di una base di norma unitaria del complemento ortogonale di un vettore gaussiano: il caso bidimensionale. STATISTICA. vol. 1, pp. 33-38 ISSN: 0390-590X.
- ◆ E.B. DAGUM, LUATI A. (2001). Asymmetric and symmetric weights of kernel smoothers and their spectral properties. ESTADISTICA. vol. 53, 160-161,

pp. 215-268 ISSN: 0014-1135.

♦ E. BEE DAGUM, LUATI A. (2003). *Fitting and smoothing properties of length constrained smoothers applied to time series*. In: M. SCHADER, W. GAUL, M. VICHI. *Between data science and applied data analysis*. (pp. 13-21). HEIDELBERG: Springer-Verlag (GERMANY).

Testo inglese

1. **DI LASCIO Francesca Marta Lilja**

Curriculum:

PRESENT OCCUPATION

PH.D. STUDENT in (Methodological) Statistics at the Department of Statistics "P. Fortunati" of the University of Bologna - Alma Mater Studiorum, Italy.
Supervisor: Prof. Paola Monari.

TEACHING ASSISTANT of the undergraduate courses "Multidimensional and Time Series Analysis" (Prof. M. Pillati) at the Faculty of Statistics, University of Bologna, Italy and "Istitution of Statistics" (Prof. A. Roverato) at the Faculty of Economy, University of Bologna, Italy.

TEACHING ASSISTANT of the graduate courses "Multivariate Statistics" (Prof. A. Montanari) at the Faculty of Statistics, University of Bologna, Italy and "Statistics for juridical-economical Applications" (Prof. A. Roverato) at the Faculty of Economy, University of Bologna, Italy.

HIGHER EDUCATION

June 1998 - *Diploma di Maturità Scientifica (GENERAL CERTIFICATE of Education in Science)*, Liceo Scientifico A.Gatto' (Agropoli, SA, Italy), with the full mark of 60/60.

September 1999 - March 2004 - *UNDERGRADUATE STUDENT* at the Faculty of Statistics of the University of Bologna, Italy.
-Passed 23 exams with the average mark of 29.5/30 with 10 laudi.

March 2004 - *Laurea in Scienze statistiche demografiche e sociali (DEGREE in Statistics)* at the Faculty of Statistics of the University of Bologna, Italy, with the mark of 110/110 CUM LAUDE.

-Degree thesis: *A new approach to the characterization of linear and non linear stochastic processes (in italian)* (supervisor: Prof. Estela Bee Dagum, University of Bologna; cosupervisor: Dott. Alessandra Luati, University of Bologna).

January 2005 - *Admission to the Ph.D. program in "(Methodological) Statistics"* at the Department of Statistics, University of Bologna, Italy.

SCIENTIFIC CURRICULUM

May 10th-14th, 2004 - *SIS SCHOOL on "Analisi delle serie storiche non lineari: teoria ed applicazioni"* at the University Campus of Fisciano, Salerno, Italy. Organized by SIS (Italian Statistical Society). Coordinator: Prof. Alessandra De Rose.

December 2004 - *First classified with the mark of 114/120 at the Public Competition for the Doctoral Program in "Methodological Statistics" of the University of Bologna (Official Gazzette n. 77, IV special series, 09.28.2004)*, Italy.

July 3rd-23rd, 2005 - *SUMMER COURSE of Statistics and Probability on "Sequential Design and Analysis with Application to Clinical Trials"*, Torgnon (AO, Italy). Organized by the IMQ (Istituto di Metodi Quantitativi) of the Bocconi University, Milan, Italy. Coordinator: Prof. Donato Michele Cifarelli.

December 9th-14th, 2005 - *5th INTERNATIONAL WINTER SCHOOL on "REASONing under PARTial Knowledge"*, Foligno (Perugia, Italy). Organized by the Mathematics and Informatics Department of the University of Perugia, Italy. Coordinator: Prof. Romano Scozzafava.

February 13th-17th, 2006 - *7th INTERNATIONAL WINTER SCHOOL on "APPLIED BIOINFORMATICS: the Test Case of Human Genome"*, Bologna, Italy. Organized by the AIRBBC (Italian Association for the Research in Biophysics and Computational Biology) c/o Department of Biology of the University of Bologna, Italy. Coordinator: Prof. Rita Casadio.

September 2006- February 2007 -- *VISITING STUDENT* at the Department of Biology, University of Pennsylvania, Philadelphia, PA, USA. Prof. Warren J. Ewens. Project's title: "Statistical Analysis of DNA sequences: dependance structure and cluster analysis in microarray experiments". Supervisor: Prof. W.J. Ewens.

TEACHING EXPERIENCES

TEACHING ASSISTANT of the undergraduate course "Multidimensional and Time Series Analysis" (Prof. M. Pillati) at the Faculty of Statistics, University of Bologna, Italy, academic year 2006/2007.

TEACHING ASSISTANT of the graduate course "Multivariate Statistics" (Prof. A. Montanari) at the Faculty of Statistics, University of Bologna, Italy, academic year 2006/2007.

RESEARCH PROJECT PARTECIPATION

April 2005 - *Pres: National Murst Project (Ex 40%) "Parametric and non parametric estimation and forecasting of time series conditional moment dynamics"*. Coordinator: Prof. Estela Bee Dagum.

SCIENTIFIC SOCIETIES MEMBERSHIP

March 2004 - *Pres: Member junior of the Italian Statistical Society (SIS)*.

October 2006 - *Pres: Member Student of the American Statistical Association (ASA)*. Sezioni di Membership: Biometrics (BIOM), Joint Stat Comp & Stat Graph (CPGH), Bayesian Statistical Science (SBSS), Teaching Stats in the Health Sciences (TSHS).

SEMINAR PRESENTATION

October 2005: *Seminar on "A new approach to the carachterization of linear and non linear stocastic processess"* held during the doctoral course on "Stochastic processess" (Prof. R. Rosa), Department of Statistics, University of Bologna, Italy.

ORGANIZED CONFERENCE

2005: Executive committee of the Final Workshop of PRIN/COFIN 2002 on "Statistical Inference on the Stochastic and Deterministic Dynamics of Observed Time Series", Bressanone, Italy, Giugno 9th-11th, 2005.

Publications:

- ◆ DI LASCIO F., EWENS WJ. (2007). Comparison between K-Means and Hierarchical Clustering of Dependent and Independent Data Generated from Multivariate Gaussian Copula Function. Annual Conference of the New Zealand Statistical Association (NZSA). 4-6 Luglio. : (NEW ZEALAND).
- ◆ DI LASCIO F. (2006). A New Starting Point to Analyze the Nonlinear Time Series: the Nonlinear ARMA Doubly Stochastic and State Dependent Models. Proceedings della 5a Conferenza Internazionale Annuale su Statistics, Mathematics and Related Field. 16 - 18 Gennaio. (pp. 453-470). : (UNITED STATES).
- ◆ DI LASCIO F. (2006). Complications of Percutaneous Laser Ablation for Hepatocellular Carcinoma: A Multicenter Study in 520 Patients. 18th Congress of Ultrasound in conjunction with the XVIII Congresso Nazionale SIUMB. 15-19 Settembre. (pp. 30). : (ITALY).
- ◆ ACCOGLI, E, TITTOTO, P, ROMANICO, A, DI LASCIO F., ET AL. (2005). Accuratezza diagnostica dell'ecografia con mezzo di contrasto (CEUS) vs l'ecografia basale (US) nella caratterizzazione dell'iperplasia nodulare focale (INF). XVII Congresso nazionale della Società Italiana di Ultrasonologia in Medicina e Biologia (SIUMB). 12 -15 Novembre. (pp. 4-5). ROMA: (ITALY).
- ◆ DI LASCIO F. (2005). Nonlinear ARMA doubly stochastic and state dependent models: a new general approach to nonlinear time series analysis. 30a Conferenza su Stochastic Processes and their Applications. 26 Giugno - 1 Luglio. (pp. 32). : (UNITED STATES).
- ◆ ROSENBERGER WF, HU F, DI LASCIO F., ET AL. (2006). Randomization for Balancing Treatment Assignments in "Sequential Design and Analysis with Application to Clinical Trials". Quaderno dell'Istituto di Metodi Quantitativi (IMQ), Università Bocconi, Milano, Italia. (vol. n. 13, parte I, cap.2., pp. 13-25).

2. **GUIDOTTI Laura**

Curriculum:

Publications:

- ◆ E. BEE DAGUM, GUIDOTTI L., A. LUATI. (2005). Some Statistical Applications of Centrosymmetric Matrices. (pp. 97 - 104). [RELAZIONE] HEIDELBERG: Springer-Verlag (GERMANY).
- ◆ GUIDOTTI L. (2005). Wilde in Arcadia. (pp. 379 - 385). ISBN/ISSN: 88 470 0346 6. [RELAZIONE] La relazione, il volume, la rassegna teatrale si configurano come attività svolte nell'ambito del progetto europeo: "Diffusion and i... MILANO: Springer- Verlag (ITALY).

3. **LUATI Alessandra**

Curriculum:

Since 2001, assistant professor at the Faculty of Statistical Sciences of the University of Bologna. Professor of Time series analysis, basic and advanced course, at the Faculty of Statistical Sciences.

Member of the Editorial Board of "Statistica", referee for various national and international journals, reviewer for Mathematical Reviews.

Member of: Italian Statistical Society, American Statistical Society, International Institute of Forecasters, Working group on Time series within the Italian Statistical Society.

Served as an active collaborator to the Research Unit of Bologna (unit and national coordinator, prof. Estela Bee Dagum) to the Miur projects granted in 2000, "Linearity and non linearity in time series dynamics", 2002, "Statistical inference on the stochastic and deterministic dynamic of observed time series" e 2004, "Parametric and non parametric estimation and forecasting of time series conditional moment dynamics".

The scientific interests concern time series analysis, with particular attention to non parametric methods for signal extraction, and statistical inference in dynamic quantum systems, with regard to information theory in such systems.

Publications:

- ◆ MANFRINI O, BAZZOCCHI G, LUATI A., BORGHI A, MONARI P, BUGIARDINI R. (2006). Coronary spasm reflects inputs from the adjacent esophageal system. AMERICAN JOURNAL OF PHYSIOLOGY. vol. 290, pp. H2085 - H2091 ISSN: 0363-6135. [ARTICOLO].
- ◆ LUATI A., G. TASSINARI. (2005). Intervention Analysis to Identify Significant Exposures in Pulsing Advertising Campaigns: an Operative Procedure. COMPUTATIONAL MANAGEMENT SCIENCE. vol. 4, n. 4, pp. 295 - 308 ISSN: 1619-697X. [ARTICOLO].
- ◆ E. BEE DAGUM, LUATI A. (2004). Relationship between local and global nonparametric estimators measures of fitting and smoothing. STUDIES IN NONLINEAR DYNAMICS AND ECONOMETRICS. vol. vol. 8 n. 2 art. 17 ISSN: 1081-1826. [ARTICOLO].
- ◆ E.B. DAGUM, LUATI A. (2004). A linear transformation and its properties with special applications in time series filtering. LINEAR ALGEBRA AND ITS APPLICATIONS. vol. 338, pp. 107 - 118 ISSN: 0024-3795. [ARTICOLO].
- ◆ LUATI A. (2004). Maximum Fisher Information in Mixed State Quantum Systems. ANNALS OF STATISTICS. vol. 32, 4, pp. 1770 - 1779 ISSN: 0090-5364. [ARTICOLO].
- ◆ LUATI A. (2003). A real formula for transition probabilities. STATISTICA. vol. 1, pp. 71-77 ISSN: 0390-590X.
- ◆ E.B. DAGUM, LUATI A. (2002). Global and local statistical properties of fixed-length nonparametric smoothers. STATISTICAL METHODS & APPLICATIONS. vol. 11, 3, pp. 313-333 ISSN: 1618-2510.
- ◆ LUATI A., P. PARUOLO. (2002). Sulla distribuzione di una base di norma unitaria del complemento ortogonale di un vettore gaussiano: il caso bidimensionale. STATISTICA. vol. 1, pp. 33-38 ISSN: 0390-590X.
- ◆ E.B. DAGUM, LUATI A. (2001). Asymmetric and symmetric weights of kernel smoothers and their spectral properties. ESTADISTICA. vol. 53, 160-161, pp. 215-268 ISSN: 0014-1135.

◆ E. BEE DAGUM, LUATI A. (2003). *Fitting and smoothing properties of length constrained smoothers applied to time series*. In: M. SCHADER, W. GAUL, M. VICHI. *Between data science and applied data analysis*. (pp. 13-21). HEIDELBERG: Springer-Verlag (GERMANY).