

**I RIUNIONE SCIENTIFICA  
DEL  
DIPARTIMENTO DI SCIENZE STATISTICHE**

**Aula 3**

**9 - 10 Febbraio 2011**

Facoltà di Ingegneria dell'Informazione,  
Informatica e Statistica  
Piazzale Aldo Moro, 5 – 00185 Roma

## MERCOLEDI' 9 FEBBRAIO 2011

### Ore 9.30 **Saluti**

Fabio Grasso, Direttore del Dipartimento di Scienze Statistiche

### Ore 9.45 **Ricordo del prof. Bruno Simeone**

Interventi di Enzo Orsingher e Gabriella Salinetti

### Ore 10.00 – 13.00

#### Moderatore: Enzo Orsingher

- **Emilia Arcaleni**, Franco Peracchi  
*Early-life environment, height and BMI of young adult males in Italy* 1
- **Luisa Beghin**  
*Fractional relaxation equations and Brownian crossing probabilities of a random boundary* 2
- **Valentina Cammarota**, Peter Mörters  
*Punti Tripli Pionieri per il Moto Browniano ed Alberi Aleatori* 5
- **Consuelo Carlevale**  
*La riassicurazione* 6
- **Oliviero Casacchia**  
*Le migrazioni interne in Italia: stranieri e italiani a confronto* 8
- **Francesco Cetta**  
*Finanza innovativa ed etica finanziaria* 10
- **Cinzia Conti**, Frank Heins, Massimo Strozza  
*La mobilità territoriale dei laureati* 11
- **Pier Luigi Conti**  
*Misurazione dell'incertezza nel matching statistico* 13
- **Alessandro De Gregorio**  
*Passeggiate aleatorie in  $R^d$  con tempi di Dirichlet* 15

# MERCOLEDI' 9 FEBBRAIO 2011

Ore 14.30 – 17.30

Moderatrice: Marcella Corsi

- **Agostino Di Ciaccio**  
*La selezione delle variabili in un modello regressivo* 17
- **Carlo D'Ippoliti**, Angela Cipollone, Marcella Corsi  
*Knowledge and Job Opportunities in a Gender Perspective: Insights from Italy* 18
- **Stefania Gubbiotti**, Pierpaolo Brutti, Fulvio De Santis  
*A robust Bayesian stopping rule for sequential trial* 20
- **Stefano Fachin**, Francesca Di Iorio  
*Testing for Cointegration with Breaks in Dependent Panels* 21
- **Paola Fersini**, Giuseppe Melisi, Valerio Scacco  
*Best Estimate e Prediction Error della Riserva Sinistri R.C.Auto in presenza dell'Indennizzo Diretto* 24
- **Salvatore Forte**, Matteo Ialenti, Marco Pirra  
*Gli effetti della dipendenza dei rischi sul pricing di una copertura congiunta incendio e furto ed r.c.auto* 27
- **Marco Pirra**, Salvatore Forte, Fabio Grasso, Matteo Ialenti  
*Solvency II: modelli interni per le riserve sinistri, tecniche di Re-reserving e Backtesting* 28
- **Massimo Franchi**, Paolo Paruolo  
*Spectral analysis of square matrix polynomials by local rank factorization* 29
- **Paolo Franciosa**, Giorgio Ausiello, Giuseppe F. Italiano, Andrea Ribichini  
*Computing Graph Spanners in Small Memory: Fault-Tolerance and Streaming* 30

# GIOVEDÌ 10 FEBBRAIO 2011

Ore 10.00 – 13.00

Moderatore: Enzo Orsingher

- **Luca Giuliano**  
*Strategie di ricerca con analisi automatica dei dati testuali* 32
- **Natalia Golini, F. Divino, Giovanna Jona Lasinio, A. Penttinen**  
*Data Augmentation Approach in Bayesian Modelling of Presence-only Data* 35
- **Mirko D'Ovidio, Enzo Orsingher**  
*Pseudo processi di Fresnel ed equazioni di ordine superiore* 36
- **Giovanna Jona Lasinio, Alan Gelfand, Mattia Jona Lasinio**  
*Analyzing spatial circular data using wrapped Gaussian processes* 39
- **Giovanna Jona Lasinio, Alan Gelfand, Mattia Jona Lasinio**  
*Data Augmentation Approach in Bayesian Modelling of Presence-only Data* 40
- **Francesco Nucci, Alberto Franco Pozzolo**  
*Exchange Rate and Wage Adjustment* 41
- **Enzo Orsingher, Valentina Cammarota**  
*Moto Browniano Iperbolico su Semispazi di Poincaré* 44
- **Maria Gabriella Ottaviani**  
*Statistica a scuola: strategie per un insegnamento efficace* 47
- **Paolo Palazzi**  
*L'economia come scienza sociale e politica* 50
- **Luca Passalacqua**  
*Sulla ottimizzazione di trattati di riassicurazione mediante tecniche di "Parallel Tempering"* 53

# GIOVEDI' 10 FEBBRAIO 2011

**Ore 14.30 – 16.30**

**Moderatore: Fabio Grasso**

- **Umberto Ferraro Petrillo**, Fabrizio Grandoni, Giuseppe F. Italiano  
*Searching with Memory Faults: An Experimental Study* 54
- **Federico Polito**, Enzo Orsingher  
*Randomly stopped nonlinear fractional birth processes* 56
- **Claudio Russo**  
*La distribuzione dei contratti di assicurazione attraverso il canale bancario* 57
- **Agostino Tripodi**, Fabio Baione, Paolo De Angelis, Andrea Fortunati  
*Longevity Risk and Financial Market: some issues* 58
- **Maurizio Vichi**  
*Research patterns, trends and new challenges in cluster analysis* 59

# **EARLY-LIFE ENVIRONMENT, HEIGHT AND BMI OF YOUNG ADULT MALES IN ITALY**

**Emilia Arcaleni**

(in collaborazione con Franco Peracchi)

We studied the relationship between early-life environment, measured by infant mortality and income per capita, and height and body-mass index (BMI) of recent cohorts of Italian males.

Living conditions during the growing years, especially in early childhood, influence body height through their impact on net nutrition, namely the balance between the supply of nutrients and the demands of metabolism, physical exertion, and disease. Thus, adult height is a useful marker of the economic and disease environment in childhood. The relative abundance of data on adult height and the link between income and height have been exploited by economic historians to analyze the well-being of populations and historical periods for which other data sources are lacking. Adult height is also a bridge between the present and the future, for it has been shown to be a strong predictor of earnings, cognitive function, and health outcomes at older ages, including longevity.

We use two types of data. For height and weight we use micro-data drawn from the military archives. These data contain information on all the Italian males who, between 1987 and 1996, took the compulsory medical examination that is part of the Army and Air Force draft process. Most of the information describing the early-life environment consists instead of aggregate data at the region- and province-level.

By combining high-quality micro-level data on height and weight with regional and province-level information, we are able to link individual height and BMI at age 18 to regional- and provincial-averages of environmental variables in the year of birth. We show that early-life environment appears to matter more for height than for BMI. Our findings are also consistent with scarring dominating selection. Finally, we find that income matters more than infant mortality, especially for height.

# Fractional relaxation equations and Brownian crossing probabilities of a random boundary

L.Beghin

*Dip. Scienze Statistiche, Sapienza Università di Roma*

La seguente equazione differenziale

$$\frac{d}{dt}p(t) = -\lambda p(t), \quad t > 0 \quad (1)$$

è nota in letteratura come *relaxation equation*. La soluzione della (1), con condizione iniziale  $p(0) = 1$ , è ovviamente pari a  $p(t) = e^{-\lambda t}$ . A partire dalla fine degli anni Novanta si è sviluppata un'ampia attività di ricerca che si è riproposta di applicare il calcolo frazionario alle equazioni della fisica matematica, sostituendo alle derivate intere di varie equazioni classiche le derivate di ordine frazionario. Anche la precedente equazione è stata estesa al caso frazionario come segue:

$$\frac{d^\nu}{dt^\nu}\psi(t) = -\lambda\psi(t), \quad t > 0 \quad (2)$$

dove  $\nu \in (0, 1)$  e  $\frac{d^\nu}{dt^\nu}$  rappresenta la derivata frazionaria nella definizione di Dzherbashyan-Caputo, ovvero

$$\frac{d^\nu}{dt^\nu}u(t) = \begin{cases} \frac{1}{\Gamma(m-\nu)} \int_0^t \frac{1}{(t-s)^{1+\nu-m}} \frac{d^m}{ds^m}u(s)ds, & \text{per } m-1 < \nu < m \\ \frac{d^m}{dt^m}u(t), & \text{per } \nu = m, \end{cases} \quad (3)$$

dove  $m = \lfloor \alpha \rfloor + 1$ . Ovviamente per  $\nu = 1$  la *fractional relaxation equation* (2) coincide con l'equazione standard (1).

L'equazione (2) è studiata in vari articoli tra i quali [8],[10] e la soluzione è data in termini analitici come funzione di Mittag-Leffler:

$$\psi_\nu(t) = E_{\nu,1}(-\lambda t^\nu), \quad (4)$$

dove

$$E_{\alpha,\beta}(z) = \sum_{r=0}^{\infty} \frac{z^r}{\Gamma(\alpha r + \beta)}, \quad \alpha, \beta \in \mathbb{C}, \quad \operatorname{Re}(\alpha), \operatorname{Re}(\beta) > 0.$$

Lo studio della fractional relaxation equation ha principalmente delle motivazioni fisiche, per spiegare le proprietà elettromagnetiche di una vasta gamma di materiali (che manifestano un decadimento a lunga memoria anziché esponenziale, cf. [14]) anche se la cosiddetta distribuzione di Mittag-Leffler trova applicazioni alla statistica (cf. ad esempio [7] e [12]) o alla teoria delle code in [13].

In realtà la soluzione  $\psi_\nu(t), t > 0$  può essere espressa in termini probabilistici in due interessanti forme. La prima è come probabilità di avere zero eventi fino al tempo  $t$  (o probabilità di sopravvivenza) per il cosiddetto *processo di Poisson frazionario*  $\mathcal{N}_\nu(t), t > 0$ , la cui legge di probabilità  $p_k^\nu(t)$  soddisfa la seguente equazione ricorsiva frazionaria

$$\frac{d^\nu p_k}{dt^\nu} = -\lambda(p_k - p_{k-1}), \quad k \geq 0, t > 0 \quad (5)$$

(cf. tra gli altri [6], [15], [9], [2], e [3]). Vale perciò la seguente uguaglianza

$$\psi_\nu(t) = p_0^\nu(t) = \Pr \{ \mathcal{N}_\nu(t) = 0 \} \quad (6)$$

e di conseguenza si possono applicare a  $\psi_\nu(t)$  i risultati ottenuti negli articoli sopra citati per il *processo di Poisson frazionario*. Ad esempio il fatto che quest'ultimo si possa esprimere come processo composto dal processo di Poisson ordinario  $N(t)$  con il tempo aleatorio  $\mathcal{T}_\nu(t)$ , ovvero  $\mathcal{N}_\nu(t) \stackrel{i.d.}{=} N(\mathcal{T}_\nu(t))$ ,  $t > 0$ . Quindi in questo caso, grazie alla (6), si ha che

$$\psi_\nu(t) = \int_0^\infty e^{-\lambda y} q_\nu(y, t) dy, \quad (7)$$

dove  $q_\nu(y, t)$  è la densità di  $\mathcal{T}_\nu$ , che a sua volta è soluzione di un'equazione di diffusione frazionaria. La (7) diventa particolarmente interessante nel caso in cui  $\nu = 1/2$ , per il quale si ha

$$\psi_{1/2}(t) = \int_0^\infty e^{-\lambda y} \frac{e^{-y^2/4t}}{\sqrt{\pi t}} dy = \Pr \{ |B(t)| < U \} = \psi(|B(t)|), \quad (8)$$

dove nell'ultima espressione  $U$  rappresenta una v.a. esponenziale di parametro  $\lambda$  e  $B$  un moto Browniano da essa indipendente. Quindi una seconda interpretazione probabilistica della soluzione alla fractional relaxation equation è quella di *crossing probability* di una barriera aleatoria da parte di un moto Browniano riflettente, nel caso  $\nu = 1/2$ . Per  $\nu = 1/2^n$  invece si potrà esprimere come

$$\psi_{1/2^n}(t) = \Pr \{ |I(t)| < U \} = \psi(|I(t)|), \quad (9)$$

ovvero come crossing probability di una barriera esponenziale da parte di un moto Browniano  $I(t)$  iterato  $(n-1)$  volte. Infine, per  $\nu = 1/3$ , la soluzione è rappresentabile in termini del processo  $A(t)$ , introdotto e studiato in [11], con funzione di trasizione

$$\bar{q}_{1/3}(y, t) = \sqrt{\frac{y}{3t}} \left[ I_{-1/3} \left( 2\sqrt{\frac{y}{33t}} \right) - I_{1/3} \left( 2\sqrt{\frac{y}{33t}} \right) \right], \quad (10)$$

dove  $I_\nu$  è la funzione di Bessel modificata. Quindi, in questo caso la soluzione della fractional relaxation equation diventa

$$\psi_{1/3}(t) = \int_0^{+\infty} e^{-\lambda y} \bar{q}_{1/3}(y, t) dy = \Pr \{ A(t) < U \}. \quad (11)$$

Si affrontano dunque varie estensioni del risultato (7) nelle seguenti direzioni:

- considerare altri tempi aleatori al posto di  $\mathcal{T}_\nu$  e quindi nella (8) al posto del moto Browniano: ad esempio, il tempo di soggiorno del moto Browniano sul semiasse positivo, il processo di Bessel, il moto Browniano elastico (in analogia all'analisi condotta in [5]);
- considerare un'altra variabile aleatoria al posto dell'esponenziale  $U$  in (8), ad esempio la Gamma;
- introdurre nella (2) una seconda derivata frazionaria o un'ipotesi di *distributed fractional derivative* (si veda [4]).

## References

- [1] **Beghin L. (2011)**, Fractional relaxation equations and Brownian crossing probabilities of a random boundary, *in preparation*.
- [2] **Beghin L., Orsingher E. (2009)**, Fractional Poisson processes and related planar random motions, *Electr. Journ. Prob.*, **14**, n.61, 1790-1826.
- [3] **Beghin L., Orsingher E. (2010)**, Poisson-type processes governed by fractional and higher-order recursive differential equations, *Electr. Journ. Prob.*, **15**, n.22, 684-709.
- [4] **Beghin L. (2010)**, Random-time processes governed by differential equations of fractional distributed order, *Technical Reports Dip. Scienze Statistiche*, n.1, submitted for publication.
- [5] **Beghin L., Orsingher E. (2010)**, Poisson process with different Brownian clocks, *under revision*.
- [6] **Laskin N. (2003)**, Fractional Poisson process, *Communications in Nonlinear Science and Numerical Simulation*, **8**, 201-213.
- [7] **Lin G. D. (1998)**, On the Mittag-Leffler distributions, *Journ. Stat. Plann. Infer.*, **74**, 1-9.
- [8] **Mainardi F. (1996)**, Fractional relaxation-oscillation and fractional diffusion-wave phenomena, *Chaos, Solitons and Fractals*, **7**, 9, 1461-1477.
- [9] **Mainardi F., Gorenflo R., Scalas E. (2004)**, A fractional generalization of the Poisson processes, *Vietnam Journ. Math.*, **32**, 53-64.
- [10] **Mainardi F., Mura A., Gorenflo R., Stojanovic M. (2007)**, The two forms of fractional relaxation of distributed order, *Journal of Vibration and Control*, **13**, 1249-1268.
- [11] **Orsingher, E., Beghin, L. (2009)**, Fractional diffusion equations and processes with randomly-varying time, *Annals of Probability*, **37** (1), 206-249.
- [12] **Pillai, R.N. (1990)**, On Mittag-Leffler functions and related distributions, *Ann. Inst. Stat. Math*, **42**, 157-161.
- [13] **Saji Kumar V.R., Pillai, R.N. (2006)**, Single server queue with batch arrivals and  $\alpha$ -Poisson distribution, *Calcutta Stat. Assoc. Bull.*, **58**, n.229-230, 93-103.
- [14] **Uchaikin V.V. (2002)**, A simple stochastic model for fractional relaxation processes, *Journ. Math. Sciences*, **111**, n.3, 3613-3622.
- [15] **Wang X.T., Zhang S.Y., Fan S. (2007)**, Non-homogeneous fractional Poisson processes, *Chaos, Solitons and Fractals*, **31**, 236-241.

# Punti Tripli Pionieri per il Moto Browniano ed Alberi Aleatori

Valentina CAMMAROTA\*      Peter MÖRTERS†

Un punto  $x \in \mathbb{R}^d$ ,  $d \geq 2$ , è un punto di molteplicità  $p$  per un moto Browniano  $\{B(t), t \geq 0\}$  in  $\mathbb{R}^d$  se esistono  $p$  tempi  $0 < t_1 < \dots < t_p$  tali che

$$x = B(t_1) = \dots = B(t_p).$$

È possibile dimostrare che quasi certamente

- se  $d \geq 4$  non esistono punti doppi,
- se  $d = 3$  esistono punti doppi, ma non esistono punti tripli,
- se  $d = 2$  esistono punti di ogni molteplicità finita.

Sia  $T$  un tempo fissato, un punto  $x \in \mathbb{R}^2$  è un punto di frontiera per  $\{B(t), 0 < t < T\}$  se è sulla frontiera della componente connessa illimitata del complemento di  $\{B(t), 0 < t < T\}$ .

Un punto  $x \in \mathbb{R}^2$  è un punto pioniero per  $\{B(t), t \geq 0\}$  se esiste un tempo  $T$  tale che  $x = B(T)$  è sulla frontiera della componente connessa illimitata del complemento di  $\{B(t), 0 < t < T\}$ .

Burdzy e Werner in [1] dimostrano che, quasi certamente, non esistono punti tripli sulla frontiera del moto Browniano piano e congetturano che non esistono punti tripli pionieri.

Seguendo l'approccio utilizzato in Kiefer e Mörters [2] si vuole dimostrare la congettura di Burdzy e Werner dimostrando l'estinzione di un albero aleatorio del tipo di Galton-Watson associato alla traiettoria del moto Browniano piano.

## References

- [1] Burdzy, K., Werner, W.: No triple point of planar Brownian motion is accessible. *Ann. Probab.*, **24**, 125-147 (1996)
- [2] Kiefer, R., Mörters, P.: The Hausdorff dimension of the double points on the Brownian frontier. *J. Theor. Probab.*, **23**, 605-623 (2010)
- [3] örters, P., Peres, Y.: Brownian motion. Cambridge Series in Statistical and Probabilistic Mathematics, vol. **30**. Cambridge University Press, Cambridge (2010)

---

\*Dipartimento di Statistica, 'SAPIENZA' UNIVERSITÀ DI ROMA, P.le A. Moro 5, 00185 Roma, Italy. *E-mail address:* [valentina.cammarota@uniroma1.it](mailto:valentina.cammarota@uniroma1.it)

†Department of Mathematical Sciences, UNIVERSITY OF BATH, Bath BA2 7AY, England. *E-mail address:* [maspm@bath.ac.uk](mailto:maspm@bath.ac.uk)

## LA RIASSICURAZIONE

Consuelo Carlevale

Fino ad un passato recente la disciplina della riassicurazione era racchiusa in pochissime norme, contenute nel codice civile (artt. 1928 e 1931 c.c.), per quanto atteneva (ed ancora attiene) ai profili meramente contrattuali; nel Testo Unico del 1959, per gli aspetti dell'organizzazione d'impresa.

Ciò tuttavia non è mai stato considerato un indice della sua marginalità. Ed anzi, diffuso e condivisibile era il convincimento della dottrina che la riassicurazione si configurasse come una fattispecie a "tipicità essenzialmente sociale", vale a dire regolata prevalentemente dagli usi e dalla prassi. Peraltro, convincimento che era giustificato sulla base della vocazione internazionale, essenzialmente dinamica, del fenomeno e dunque della difficoltà intrinseca di ricondurlo ad un sistema normativo rigido e preordinato.

La ricerca prende le mosse dal constatare come, negli ultimi anni, l'approccio del legislatore al tema si sia completamente modificato.

L'impressione è che si sia voluto valorizzare – attraverso una loro progressiva implementazione – il ruolo delle fonti legali su quelle, in senso lato, convenzionali. Ed in questo senso va esaminato il testo della Direttiva 2005/68/CE, attuata in Italia con D.lgs. 29 febbraio 2008, n. 56, che ha a sua volta modificato ed integrato le disposizioni del codice delle assicurazioni relative alla riassicurazione (artt. 57 e 67 cod. ass.).

La questione è quali effetti, sia in termini di ricostruzione giuridica del fenomeno sia di congruità della disciplina rispetto alle esigenze concrete del mercato, ne siano derivati.

Procedendo per punti, come del resto suggerisce la natura necessariamente sintetica delle presenti considerazioni, mi sembra che la novella incida in primo luogo sul risalente dibattito, relativo alla natura giuridica della fattispecie riassicurazione.

Si discute, in proposito, se si tratti di un contratto di assicurazione – il dubbio discende dal fatto che non rientra nella definizione della norma capostipite (art. 1882 c.c.) – e se, risolto positivamente il primo quesito, esso segua la disciplina delle polizze sulla vita ovvero sui danni, che formano la massa dei rischi riassicurati.

Ebbene, allo stato della ricerca, fermo che l'impianto codicistico non ha subito variazioni, la professionalità delle parti del contratto di riassicurazione esclude l'applicabilità di tutta la disciplina sui consumatori, vale la pena registrare un allargamento della fattispecie a contratti a connotazione, in realtà, essenzialmente finanziaria.

Mi riferisco, in particolare, alla *finite reinsurance*. Il contratto prevede che il riassicuratore limiti la propria esposizione ("*finite*") fissando un massimale, che deve essere superiore al premio incassato per un importo almeno pari al rendimento che il riassicuratore prevede di ricavare dall'investimento del premio quando, dopo qualche anno (il contratto ha sempre una durata pluriennale), dovrà pagare il massimo convenuto. Ed ancora alle c.d. società veicolo, la cui attività consiste nel collocamento presso il pubblico dei risparmiatori di

strumenti finanziari garantiti dall'esercizio e quindi dall'andamento dell'attività riassicurativa, anch'essa normalmente spalmata su un periodo pluriennale.

Da ciò, pare ragionevole desumere che la riassicurazione sia configurabile, ormai pienamente, come una fattispecie a sé. Circostanza da cui consegue che ad essa si applica esclusivamente la disciplina che le è propria, mentre risulterebbe probabilmente sterile il tentativo di integrarla con il ricorso alle norme ora sulle polizze vita, ora su quelle danni, come invece prospettato dalla dottrina più risalente.

D'altro canto, l'esistenza di una definizione normativa, per quanto lata, espone alla necessità di verificare se la proliferazioni di clausole convenzionali, fino ora frequentissima sul mercato, sia ancora (ed entro quali limiti) connotata da profili di legittimità.

Per quanto concerne l'incidenza "pratica" della novella, in termini di funzionalità operativa, il presente contesto consente di fare cenno soltanto ad uno dei problemi applicativi che potranno discenderne.

Fino all'entrata in vigore del DLgs. N. 56/2008, le imprese di riassicurazione italiane non avevano bisogno di alcuna autorizzazione per operare in regime di libera prestazione di servizi. Ad oggi, invece (v. art. 59 quater cod. ass.) occorre comunicare l'intento ad Isvap, precisandone le modalità. Il processo presenta significativi dubbi di conformità al principio dell'autorizzazione unica e sembra trascurare importanti elementi tecnico – operativi delle contrattazioni riassicurative, dal momento che le relazioni fra riassicuratore e riassicurato tollerano scarsamente restrizioni operative e si muovono, praticamente sempre, in contesti internazionali.

# Le migrazioni interne in Italia: stranieri e italiani a confronto

Oliviero Casacchia

## Obiettivi

In Italia, tra il 1995 e il 2008, i trasferimenti di residenza tra comuni italiani sono aumentati di oltre il 25%. Si è dunque in presenza di un periodo di ripresa della mobilità che pone tale fenomeno all'attenzione di molti autori (Livi Bacci 2007, Bonifazi 2009, De Santis, 2010). Nel nostro paese al 2008 sono stati registrati circa un milione e 389 mila cambiamenti di residenza, dato che appare in crescita sia rispetto all'anno precedente (+ 9 mila) sia rispetto al 1995 (+25%). Un aspetto rilevante è il contributo degli stranieri che nell'ultimo anno considerato sono arrivati a comporre il 15% del totale dei cambiamenti di residenza. L'aumento del numero di cittadini stranieri residenti, la loro maggiore propensione a spostarsi all'interno del Paese e i fattori strutturali spiegano parzialmente il fenomeno. Contemporaneamente anche la struttura delle direttrici appare in parte diversa da quella degli italiani, a testimoniare l'esistenza di percorsi privilegiati e fattori di attrazione che agiscono in modo differente sulla popolazione straniera.

L'obiettivo della ricerca, condotta con altri colleghi di Università italiane (Napoli, Roma Tre e Teramo) e con l'Istat, è di individuare le principali direttrici e le caratteristiche strutturali delle migrazioni interne degli stranieri e degli italiani. Ci si propone di evidenziare i fattori attrazione che giocano un ruolo differente tra italiani e stranieri. Accanto alle misure tradizionali, viene utilizzato un modello gravitazionale che quantifica gli effetti sulla mobilità interna della consistenza delle popolazioni di origine e di destinazione dei flussi e della distanza tra le aree territoriali.

## Dati e metodi

La fonte utilizzata è quella anagrafica, la documentazione maggiormente utilizzata negli studi sulla mobilità interna. L'attenzione della ricerca è posta sulla mobilità a medio e lungo raggio, quella tra province, in quanto sebbene la mobilità intraprovinciale rappresenti in Italia la principale tipologia di spostamento essa esprime soprattutto una mobilità di breve raggio che presenta caratteristiche del tutto proprie e non assimilabili a quelle delle altre tipologie. La mobilità a medio e lungo raggio è quella che ha un maggior impatto demografico, economico e sociale: si è scelto così di considerare gli spostamenti che attraversano almeno il confine della provincia. Il tema della mobilità a breve raggio è stato affrontato ricorrendo ad un'analisi della mobilità intraurbana, prendendo come *case study* il comune di Roma.

La prima parte del lavoro si concentra sull'analisi descrittiva della mobilità interna distinguendo il comportamento degli italiani e degli stranieri, cogliendone similitudini e dissimilarità. Successivamente si è scelto di applicare il modello gravitazionale ai flussi migratori tra le province secondo l'ipotesi che l'ammontare dei flussi tra due zone sia direttamente proporzionale alla *consistenza* delle popolazioni di origine e di destinazione e inversamente proporzionale alla *distanza* tra le due zone considerate. Lo stesso modello è stato poi applicato con riferimento alla mobilità interna a Roma, considerando sia i 19 municipi che aggregazioni specifiche formate dalle macro-zone.

## Risultati

Per i cittadini stranieri si colgono livelli di mobilità nettamente più elevati rispetto agli italiani: sono collettivi più mobili probabilmente perché gli stranieri hanno meno vincoli (come quelli legati alla formazione di una famiglia, la sistemazione residenziale, e così via) e caratteristiche strutturali che favoriscono la mobilità come la prevalenza delle età giovanili, periodo nel quale in cui ci si sposta più frequentemente.

Il modello gravitazionale ha consentito di confermare come le tre variabili macro rappresentino un'importante componente, da non trascurare, nello studio del fenomeno migratorio. Inoltre i modelli hanno permesso di evidenziare come l'ammontare delle popolazioni di origine e di destinazione e la distanza compiuta negli spostamenti abbiano un effetto differente nelle applicazioni effettuate per interpretare la mobilità interna di italiani e stranieri. In particolare, mentre nei modelli interprovinciali emerge sorprendentemente un effetto della variabile distanza più consistente nel caso della mobilità degli stranieri, l'inverso si verifica per quanto concerne la mobilità all'interno di Roma.

## Riferimenti bibliografici

- Bonifazi C. (2009) Le migrazioni interne meridionali: vecchi e nuovi ritardi, in *www.neodemos.it*
- De Santis (2010) Mobilità a corto e a lungo raggio e pendolarismo della popolazione italiana, in *Demografia del capitale umano* Livi Bacci M. (a cura di), il Mulino.
- Livi Bacci (2007) Ma c'e' davvero una ripresa delle migrazioni sud-nord?, in *www.neodemos.it*
- Poulain M. (1981) Entre la gravitation et l'entropie: un modèle spatiale de la migration interne et son estimation, comunicazione alla *tavola rotonda sulla Demografia Matematica*, Unione internazionale per lo studio della popolazione (IUSSP), Manila.

## **FINANZA INNOVATIVA ED ETICA FINANZIARIA**

**Francesco Cetta**

Se ne parla da alcuni anni con effetti contrapposti sui mercati internazionali ormai globalizzati. Mi chiedo dove sia il “border line” di questo processo evolutivo (oppure involutivo?). La crisi finanziaria ha indubbiamente prodotto un “effetto domino” dal quale non si può essere immunizzati del tutto, vediamo ad esempio le oscillazioni borsistiche e le fluttuazioni dei tassi di cambio, in particolare euro e dollaro americano. Le cause specifiche sono state addebitate all’uso e all’abuso della cosiddetta “finanza innovativa” con la spregiudicatezza del trading di strumenti finanziari derivati che meno indirizzati alla copertura di rischi finanziari si sono per lo più rivolti alle azioni speculative degli intermediari. Allora gli “hedge fund” hanno avuto modo di proliferare con tutte le conseguenze economiche denunciate dagli stati sovrani.

Forse è giunto il momento di una presa di coscienza immediata che si riferisca all’etica finanziaria degli operatori, i quali nella logica della massimizzazione del profitto non dimentichino le aspettative degli investitori.

Ritengo che solo in questo modo si possano sanare tante situazioni a limite di default. La ricerca scientifica nel settore economico-finanziario ha fornito negli anni significativi risultati (ad esempio i modelli matematici sull’informazione asimmetrica e la teoria dei contratti).

Invito pertanto i giovani ricercatori del Dipartimento di Scienze Statistiche ad implementare tali linee di indagine.

# La mobilità territoriale dei laureati

Cinzia Conti<sup>1</sup>, Frank Heins<sup>2</sup>, Massimo Strozza<sup>1</sup>,

<sup>1</sup> ISTAT, Direzione Centrale per le Statistiche e le Indagini sulle Istituzioni Sociali

<sup>2</sup> Istituto di Ricerche sulla Popolazione e le Politiche Sociali, CNR, Roma  
e-mail: ciconti@istat.it

## Obiettivi/Aims

Nell'ambito della mobilità interna, gli spostamenti sul territorio di giovani *high skilled* rappresentano un aspetto di particolare interesse per le conseguenze sul capitale umano sia delle aree di provenienza sia di quelle di arrivo.

La mobilità territoriale talvolta si configura come una scelta finalizzata a perseguire nella maniera migliore i propri obiettivi formativi o lavorativi. In altri casi, invece, si configura come una necessità determinata dalla non adeguata offerta formativa o da un mercato del lavoro che offre limitate o poco qualificate opportunità di impiego. Al di là del contesto economico-sociale di "origine", anche i tratti demografici dei giovani adulti e il loro background familiare possono determinare comportamenti e percorsi differenti. Il lavoro proposto si pone come obiettivo di analisi quello di individuare, sulla base delle caratteristiche sopra specificate, differenti profili di giovani che si spostano per studio o per lavoro all'interno del nostro Paese.

## Dati e metodi / Data and methods

Il contributo parte dall'analisi dei dati sui trasferimenti di residenza riguardanti i giovani tra 20 e 34 anni. L'approfondimento verrà svolto attraverso i dati delle ultime edizioni delle indagini campionarie Istat sulla transizione istruzione – lavoro.

L'analisi descrittiva sarà seguita da analisi condotte attraverso modelli di regressione logistica multivariate al fine di cogliere eventuali relazioni tra la propensione alla mobilità e le caratteristiche dei giovani laureati che, per scelta o per necessità, danno vita a comportamenti differenti.

Nell'analisi si presterà attenzione alle distanze osservate fra provincia d'origine, provincia di studio e provincia di residenza attuale.

## Risultati/Results

La mobilità dei giovani diplomati e laureati segue delle direttrici peculiari rispetto ai trasferimenti di residenza che, in generale, avvengono in Italia.

Gli spostamenti per studio rappresentano una quota non irrilevante della mobilità interna.

Non deve essere trascurato che i movimenti migratori che avvengono per motivi formativi possono non essere transitori e dare vita, nel più lungo periodo, a trasferimenti di residenza "definitivi".

Ciò comporta che i territori sedi di Atenei importanti esercitino un potere attrattivo non solo sui giovanissimi, ma anche su coloro che, terminati gli studi, possono decidere per molteplici motivazioni (formazione della famiglia, lavoro, etc.) di restare nel luogo dove hanno studiato.

I profili dei gruppi di giovani individuati mettono in luce come le differenti caratteristiche demografiche e socio economico si associno a differenti comportamenti di mobilità. L'analisi dei profili consente anche di discriminare tra scelta di mobilità e necessità di spostarsi.

Infine l'analisi conferma la rilevanza del territorio, come variabile *proxy* esplicativa dei comportamenti seguiti.

## Riferimenti bibliografici/References

- Brait F., Petrillo R., Strozza M. (2010) Student and Graduate Mobility in Italy. *Bulletin of Comparative Labour Relations*, Volume 73, 149-179
- D'Antonio M., Scarlato M. (2007) *I laureati del Mezzogiorno: una risorsa sottoutilizzata o dispersa*. Quaderno Svimez 10, SVIMEZ, Roma.
- Dal Bianco A. (2007) *Determinants of Student Migration in Italy*, <[www.internet.it/aisre/minisito\\_2007/cd\\_aisre/index.html](http://www.internet.it/aisre/minisito_2007/cd_aisre/index.html)>.
- De Santis G. (2010) Mobilità a corto e lungo raggio e pendolarismo della popolazione italiana, in Livi Bacci M. (a cura di) *Demografia del capitale umano*, Il Mulino, Bologna, 123-138.
- Rossi F., Strozza S. (2007) Mobilità della popolazione, immigrazione e presenza straniera, in GCD-SIS, *Rapporto sulla popolazione. L'Italia all'inizio del XXI secolo*, il Mulino, Bologna, 111-137.
- Viesti G. (2005) Nuove migrazioni. Il trasferimento di forza lavoro giovane e qualificata dal Sud al Nord, *Rivista Bimestrale di cultura e di politica*, Il Mulino, 4, 678-688

## Misurazione dell'incertezza nel matching statistico

Pier Luigi Conti

Dipartimento di Scienze Statistiche - Sapienza Università di Roma

Il *matching* statistico (detto talvolta “fusione dei dati”) consiste nel combinare le informazioni provenienti da differenti rilevazioni campionarie, in ciascuna delle quali si osservano solo alcune delle variabili statistiche di interesse. L'obiettivo ultimo, in buona sostanza, è quello di produrre un unico *file* di dati sintetico, contenente tutte le variabili di interesse.

Per essere più concreti, si considererà il caso paradigmatico in cui l'interesse verte su tre variabili statistiche, indicate rispettivamente con  $X$ ,  $Y$ ,  $Z$ . Siano poi  $A$  e  $B$  due campioni indipendenti, rispettivamente di numerosità  $n_A$  e  $n_B$ . A causa del processo di osservazione, solo  $(X, Y)$  are observed in  $A$ , e solo  $(X, Z)$  sono osservate in  $B$ . Quindi, la v.a.  $Z$  ( $Y$ ) è mancante in  $A$  ( $B$ ).

In simboli, le osservazioni campionarie disponibili sono indicate con

- Campione  $A$ :  $(X_1^A, Y_1^A), \dots, (X_{n_A}^A, Y_{n_A}^A)$ ;
- Campione  $B$ :  $(X_1^B, Z_1^B), \dots, (X_{n_B}^B, Z_{n_B}^B)$ .

Come già accennato, *il matching statistico consiste nel produrre un insieme di dati “completo”, in cui le tre variabili di interesse  $X$ ,  $Y$ ,  $Z$  sono simultaneamente presenti.*

La conseguenza principale del fatto che non sono disponibili osservazioni simultanee di  $X$ ,  $Y$ ,  $Z$  è che *il meccanismo di osservazione campionaria non è in grado di identificare la distribuzione congiunta di  $(X, Y, Z)$ .*

In letteratura due sono le tecniche principali per affrontare questo problema.

1. Tecniche basate sull'assunzione di indipendenza condizionata (CIA) tra  $Y$  e  $Z$  dato  $X$ ;
2. Tecniche basate su informazioni ausiliarie, come ad es., un “piccolo” campione  $C$  contenente osservazioni congiunte di  $(X, Z, Y)$ .

Le tecniche copera menzionate rendono la distribuzione congiunta di  $(X, Y, Z)$  identificabile da parte del meccanismo di osservazione campionario. Qualora per la CIA non possa essere ragionevolmente assunta (come in genere accade in pratica) o non siano disponibili informazioni

ausiliarie, la distribuzione congiunta di  $(X, Z, Y)$  è non identificabile. In questo caso si ha *incertezza* sul modello per  $(X, Y, Z)$ .

Gli obiettivi principali della presentazione sono di seguito elencati.

1. Studio della nozione di incertezza relativa al modello statistico in ambito non parametrico.
2. Definizione di una misura di incertezza per modelli non identificabili.
3. Studio dell'effetto sulla misura dell'incertezza dovuto a vincoli logici sulle relazioni tra  $Y$  e  $Z$  dato  $X$ .
4. Definizione di uno stimatore della misura di incertezza proposta, e studio delle sue proprietà.

# Passeggiate aleatorie in $\mathbb{R}^d$ con tempi di Dirichlet

Alessandro De Gregorio  
 Dipartimento di Scienze Statistiche  
 "Sapienza" University of Rome  
 P.le Aldo Moro, 5 - 00185, Rome - Italy  
 alessandro.degregorio@uniroma1.it

Consideriamo delle passeggiate aleatorie a tempo continuo in  $\mathbb{R}^d$ ,  $d \geq 2$ , definite mediante una particella che cambia direzione un numero fissato di volte  $n \geq 1$  e con velocità costante  $c > 0$ . Le direzioni sono scelte uniformemente su un'ipersfera di raggio unitario ed in modo indipendente. Assumiamo che gli intervalli di tempo tra due istanti consecutivi nei quali si è verificato il cambio direzione, abbiano distribuzione congiunta data da una Dirichlet. Questo processo aleatorio, originariamente introdotto da Pearson e anche noto in letteratura con il nome di *volo aleatorio*, sembra essere molto utile nel descrivere moti reali. In effetti i voli aleatori hanno trovato applicazione in svariati campi della scienza: fisica, biologia, astronomia.

In questo lavoro si dimostra che la funzione caratteristica della posizione raggiunta dalla particella  $\underline{\mathbf{x}}_d(t)$ ,  $t > 0$ , dati  $n + 1$  passi, è esprimibile mediante funzioni di Bessel, ovvero

$$E \left\{ e^{i \langle \underline{\alpha}_d, \underline{\mathbf{x}}_d(t) \rangle} \right\} = \frac{2^{\frac{n+1}{2}(d-1) - \frac{1}{2}} \Gamma\left(\frac{n+1}{2}(d-1) + \frac{1}{2}\right)}{(ct \|\underline{\alpha}_d\|)^{\frac{n+1}{2}(d-1) - \frac{1}{2}}} J_{\frac{n+1}{2}(d-1) - \frac{1}{2}}(ct \|\underline{\alpha}_d\|),$$

dove  $J_\mu(x) = \sum_{k=0}^{\infty} (-1)^k \frac{(x/2)^{2k+\mu}}{k! \Gamma(k+\mu+1)}$ ,  $x, \mu \in \mathbb{R}$ . Tutto ciò permette di ottenere in modo esplicito le distribuzioni di probabilità invertendo la suddetta funzione caratteristica. Pertanto si ha che la funzione di densità è uguale a

$$p_{\underline{\mathbf{x}}_d}(\underline{\mathbf{x}}_d, t; n) = \frac{\Gamma\left(\frac{n+1}{2}(d-1) + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}(d-1)\right)} \frac{(c^2 t^2 - \|\underline{\mathbf{x}}_d\|^2)^{\frac{n}{2}(d-1) - 1}}{\pi^{d/2} (ct)^{(n+1)(d-1) - 1}},$$

con  $d \geq 2$ ,  $\|\underline{\mathbf{x}}_d\| < ct$  e  $n \geq 1$ . È interessante confrontare questi risultati con quelli ottenuti nel caso standard dove i tempi sono considerati uniformi sul semplice.

Se il numero di cambi di direzione è randomizzato, per mezzo di un processo di Poisson frazionario siamo in grado di ottenere la distribuzione non condizionata  $P\{\underline{\mathbf{x}}_d(t) \in d\underline{\mathbf{x}}_d\}$  per ogni  $d \geq 2$ .

Discuteremo anche la possibilità di considerare direzioni distribuite con legge di probabilità diversa da quella uniforme. La particella sotto tale ipotesi tenderebbe a muoversi su alcune regioni dello spazio con maggiore probabilità; in altri termini si avrebbe un moto con deriva. Ovviamente questa ipotesi determina una serie di conseguenze sul modello quali la perdita della isotropia e della ricorrenza. Questo fa sì che dal punto di vista analitico il moto aleatorio con deriva sia molto più complicato da trattare.

## BIBLIOGRAFIA

Beghin, L., Orsingher, E. (2010) Moving randomly amid scattered obstacles. *Stochastics*, **82**, 201-229.

De Gregorio, A., Orsingher, E. (2010) Flying randomly in  $\mathbb{R}^d$  with Dirichlet displacements. *Rapporto tecnico*.

Francheschetti, M. (2007) When a random walk of fixed length can lead uniformly anywhere inside a hypersphere. *Journal of Statistical Physics*, **127**, 813-823.

Garcia-Pelayo, R. (2008) Twice scattered particles in a plane are uniformly distributed. *Journal of Statistical Physics*, **133**, 401-404.

Gradshteyn I.S., Ryzhik I.M. (1980) *Table of integrals, series and products*. Academic Press, New York.

Hughes, B.D. (1995) *Random walks and Random Environments. Volume 1: Random walks*. Oxford Science Publications.

Kluyver, J.C. (1905) A local probability problem. *Proceedings of the Section of Sciences, Koninklijke Akademie van Wetenschappen te Amsterdam*, **8**, 341-350.

Le Caer, G. (2010) A Pearson-Dirichlet random walk, *Journal of Statistical Physics*, **140**, 728-751.

Masoliver M., Porrá J.M., Weiss G.H. (1993) Some two and three-dimensional persistent random walk. *Physica A*, **193**, 469-482.

Orsingher, E., De Gregorio, A. (2007) Random flights in higher spaces. *Journal of Theoretical Probability*, **20**, 769-806.

Rayleigh, L. (1919) On the problem of the random flights and of random vibrations in one, two and three dimensions. *Philosophical Magazine*, **37**, 321-347.

Stadje W. (1987) The exact probability distribution of a two-dimensional random walk. *Journal of Statistical Physics*, **46**, 207-216.

## La selezione delle variabili in un modello regressivo

Agostino Di Ciaccio

Se siamo interessati alla relazione che lega una variabile quantitativa  $Y$  a  $J$  variabili quantitative  $X_j$ , possiamo costruire un modello di regressione della  $Y$  in funzione delle  $X_j$ , tenendo conto che una parte di queste variabili esplicative può essere inutile, se non di disturbo, alla nostra analisi. In tal caso è opportuno mettere in atto una qualche procedura di selezione delle variabili che ci permetta di individuare quelle effettivamente “rilevanti”, eliminando dal modello tutte le altre. La presenza di variabili di disturbo può infatti ridurre in maniera importante la performance del modello.

A questo punto però si pone un problema di scelta tra due diversi scenari:

- 1) Supponiamo che esista il “vero modello” e che la classe di modelli che stiamo considerando lo contiene. Siamo quindi alla ricerca del “vero modello”, che dovrà includere tutte e sole le variabili importanti per spiegare la  $Y$ .
- 2) Non supponiamo l’esistenza di un “modello vero” e l’aspetto interpretativo è secondario. Siamo quindi alla ricerca di un modello efficace, in particolare dal punto di vista predittivo.

Se non siamo alla ricerca del “vero modello”, la definizione di cosa sia una variabile “importante” non è ovvia, anche se è prevalente l’interpretazione che una variabile è importante se la sua eliminazione peggiora in modo rilevante l’accuratezza predittiva del modello (Breiman 2001). D’altra parte il significato stesso di “vero modello” è stato criticato da molti autori (ad es. Akaike 1985, Shao 1997).

E’ importante sottolineare che per dimensioni campionarie realistiche, i due scenari possono portare all’individuazione di modelli molto diversi.

C’è un’ampia letteratura su questa problematica e continuano ad apparire contributi importanti sulle principali riviste, rendendo questo tema particolarmente attuale e stimolante.

Ricordiamo alcune proposte presenti in letteratura:

AIC (Akaike 1973), Cp (Mallows 1973), CV-1 (Allen 1974), CV-d (Geisser 1975, Shao 1993), BIC (Scharz 1978), GCV (Craven & Wahba 1979), FPE (Shibata 1984), GIC (Nishii 1984), RIC (Foster and George 1994), LASSO (Tibshirani 1996), SCAD (Fan & Li 2001), LARS (Efron et al. 2004), Relaxed Lasso (Meinshausen 2007), MC+ (Zhang 2010), Intrinsic Bayes Factor (Casella & Moreno 2006), Model Averaging (Hjort & Claeskens 2003).

Ricordiamo anche alcuni contributi specifici in presenza di grandi matrici di dati:

Ma & Zhang (2008), Fan & Lv (2009), Fan & Li(2006), Greenshtein (2006), Zhang & Huang (2008), Wasserman & Roeder (2009), Huang et al.(2010).

Alla luce di tutte queste proposte, che considerano dati a volte poco realistici e proprietà non sempre comparabili, è difficile capire quale criterio scegliere nelle situazioni reali.

In letteratura è stata studiata per lo più la consistenza di alcuni criteri, fornendo definizioni a volte diverse e effettuando simulazioni su pochi casi. In genere non si tiene conto di alcune caratteristiche dei dati che hanno certamente rilevanza nella performance di un criterio: numerosità campionaria fissata, rapporto segnale/rumore, numero di variabili analizzate, livello di correlazione tra le variabili, congruità della classe di modelli scelta.

A questo si potrebbe ancora aggiungere che per un numero di variabili non piccolo, è necessario scegliere anche la strategia di ricerca (non esaustiva) del modello ottimale.

In questa comunicazione analizzeremo le caratteristiche di alcuni criteri di selezione, confrontando le loro performance attraverso delle ampie simulazioni e mettendo in evidenza le caratteristiche dei dati che assumono rilevanza per la scelta del criterio.

## Knowledge and Job Opportunities in a Gender Perspective: Insights from Italy

Angela Cipollone, Marcella Corsi, Carlo D'Ippoliti <sup>1</sup>

Il recente dibattito sull'accesso ai sistemi di istruzione e sull'impatto dell'istruzione e della formazione sulle prospettive di partecipazione al mercato del lavoro ha condotto a due risultati ampiamente condivisi. Da un lato, un certo numero di analisi guarda alle opportunità di accesso ad istruzione e formazione di qualità: molte hanno evidenziato come il capitale sociale sia la variabile determinante nell'accesso all'istruzione, soprattutto all'istruzione secondaria superiore e all'istruzione universitaria. Le probabilità di ottenere titoli di studio oltre l'obbligo normativo e di una buona performance scolastica sono tanto maggiori quanto più il background familiare è istruito e benestante. Questo fenomeno di selezione è un primo indicatore del grado di disuguaglianza nell'accesso all'istruzione e alla formazione.

Un secondo gruppo di analisi guarda all'impatto del livello di istruzione conseguito sulle prospettive di partecipazione al mercato del lavoro e sulla dinamica dei redditi da lavoro. Tale impatto è significativamente positivo: l'istruzione e la formazione garantiscono una più stabile partecipazione al mercato del lavoro e più alte potenzialità di reddito da lavoro.

Entrambi questi filoni di ricerca sono stati ampiamente studiati in ottica di genere. Nel corso degli ultimi anni si è assistito ad una progressiva crescita del livello di istruzione delle donne, che tuttavia si accompagna al persistere di maggiori difficoltà nel grado di partecipazione al mercato del lavoro rispetto alla componente maschile. Il possesso di un elevato livello di istruzione, infatti, è in grado di agevolare sensibilmente la partecipazione delle donne al mercato del lavoro e di ridurre in modo significativo il divario di genere nel tasso di attività, mentre per gli uomini il possesso di un titolo di studio debole è statisticamente meno discriminante. Lo stesso risultato si ottiene guardando ad istruzione e formazione come determinanti del salario: le donne ottengono generalmente più alti rendimenti dall'istruzione, mentre gli uomini ottengono rendimenti più alti dalla posizione occupazionale. Questi risultati hanno poi condotto all'avanzamento di numerose ipotesi e stime sui differenziali di genere che distinguono tra differenziale retributivo da attribuire alle differenti caratteristiche della forza lavoro maschile e femminile e differenziale retributivo da attribuire a fenomeni di discriminazione.

Obiettivo del presente lavoro è ripercorrere le dinamiche indicate e proporre una nuova, più estesa, misura delle competenze individuali: una misura di conoscenza (knowledge). Risulta infatti interessante valutare l'impatto di tale misura, anche relativamente a quella tipicamente legata solo all'istruzione, sulle opportunità di impiego e di reddito in ottica di genere.

Una misura di conoscenza offre il vantaggio di valutare l'impatto congiunto delle variabili che determinano le competenze individuali (quali, ad esempio, istruzione, formazione, esperienza specifica sul posto di lavoro, esperienza sul mercato del lavoro) sulla partecipazione al mercato del lavoro e sulla dinamica dei salari, nell'ipotesi che il mercato del lavoro non remunererà singolarmente ciascuna competenza, ma crei un mercato per competenze congiunte, caratteristiche individuali non osservabili e per gestione del tempo.

---

<sup>1</sup> Paper prepared for the ASSA meeting 2011, IAFFE Session "Frontiers of Feminist Economics"

L'esercizio empirico proposto vede dunque l'analisi della conoscenza sulle opportunità di impiego e sul reddito di uomini e donne attraverso l'utilizzo dei dati prodotti dalla Banca d'Italia (Indagine sui Bilanci delle Famiglie Italiane) per il 2004. In tale anno, infatti, una sezione speciale del questionario riguardava diverse competenze e conoscenze degli intervistati.

Tra i principali risultati vi è l'evidenza che le singole competenze individuali interagiscono in maniera eterogenea tra gli individui, in funzione della qualità e tipologia dei sistemi educativi, della regione di residenza, dell'abilità individuale, del background familiare, e dello sforzo e del tempo che si esercitano nelle attività di istruzione e formazione e nella partecipazione al mercato del lavoro. Questa considerazione diventa ancora più cruciale in un'ottica di genere, in quanto in Italia l'interazione tra uso del tempo, regione di residenza, background familiare, tipologia di istruzione e partecipazione al lavoro è sensibilmente diversa tra uomini e donne, creando così un diverso impatto congiunto di skill e istruzione (conoscenza) sulle opportunità di impiego di uomini e donne. In particolare, emerge dall'analisi che le donne beneficiano di maggiori rendimenti della conoscenza, in termini di occupabilità e di reddito da lavoro. Però, nonostante nei livelli di istruzione il divario di genere sia ormai colmato, le donne ancora presentano un più basso livello generale di conoscenza, principalmente a causa della loro minore accumulazione di conoscenza legata al mercato del lavoro. Le conoscenze tacite acquisite mediante il lavoro non retribuito non sono adeguatamente riconosciute dal mercato, ed anzi impegni di cura di bambini e anziani sono ancora la principale barriera all'occupazione femminile.

## A robust Bayesian stopping rule for sequential trials

*Pierpaolo Brutti, Fulvio De Santis, Stefania Gubbiotti*

### Abstract

This work deals with sequential clinical trials. We denote by  $\theta$  the parameter of interest representing treatment effect. In a Bayesian context, we start formalizing pre-experimental information through a prior probability distribution on  $\theta$ . By iteratively applying Bayes theorem, the posterior distribution is derived after each observation is collected. Hence we monitor the posterior probability that  $\theta$  exceeds a minimally clinical relevant threshold as the sample size sequentially increases. Then the trial is terminated with success when it is larger than a given cutoff; otherwise, the treatment is declared ineffective. In this setting the sample size is a random variable associated to the chosen stopping rule. We show by simulation that its expectation is smaller than the non sequential optimal sample size. Moreover, we consider the issue of robustness with respect to the prior specification. We define a robust Bayesian stopping criterion that is, in general, more conservative than the non robust one. However, we show that, working sequentially, we can save observations even though a robust stopping rule is considered. More precisely, we evaluate the critical level of robustness we can afford to have a sample size that reaches the non sequential non robust optimal one.

### References

- [1] Shih, W.J., Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: A comparison. *Statist. Med.* **25**, 933–941 (2006) doi: 10.1002/sim.2252
- [2] Spiegelhalter D. J, Abrams K. R. and Myles J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. Wiley.
- [3] De Santis, F. Sample size determination for robust Bayesian analysis. *Journal of the American Statistical Association*, 2006; **101**, n. 473, 278-291.
- [4] Brutti P., De Santis F., Gubbiotti S. Robust Bayesian sample size determination in clinical trials. *Statistics in Medicine*, 2008; **27**, 2290-2306
- [5] Brutti P., De Santis F., Gubbiotti S. Mixtures of prior distributions for predictive Bayesian sample size calculations in clinical trials. *Statistics in Medicine*, 2009; **28**, 2185-2201

## Testing for Cointegration with Breaks in Dependent Panels

Francesca Di Iorio  
*University of Naples Federico II*

Stefano Fachin  
*University of Rome "La Sapienza"*

Allowing for structural breaks in the coefficients is a natural generalisation for any test of the existence of a relationship between a set of variables. In this vein Gregory and Hansen (1996) proposed a test for cointegration with breaks which, however, tends to have rather poor power in small or moderate samples. A further natural step forward is thus to construct a panel cointegration test with breaks, which can be expected to grant considerable power gains over pure time series procedures. Unfortunately, the empirical appeal of the currently available panel cointegration test with breaks is severely limited by their restrictive assumptions on the dependence structure of the panel: either full cross-section independence (Gutierrez, 2009, Westerlund, 2006a,b) or special types common factors (Banerjee and Carrion-i-Silvestre, 2006, Westerlund and Edgerton, 2008). We must thus conclude that a panel cointegration test with breaks and full cross-section dependence is not available yet. We tackle the problem extending to the case of unknown breaks the panel cointegration test proposed in Di Iorio and Fachin (2010a).

Let us consider for simplicity a standard bivariate panel cointegration set-up, with interest variables, denoted as usual by  $X$  and  $Y$ , observed over  $N$  units and  $T$  time periods. In each unit  $X$  and  $Y$  are linked by a linear, not necessarily cointegrating, relationship with a break in period  $t_i^b$ :

$$y_{it} = \begin{cases} \mu_{0i} + \beta_0 x_{it} + \epsilon_{it}, & t \leq t_i^b \\ \mu_{1i} + \beta_1 x_{it} + \epsilon_{it}, & t > t_i^b \end{cases} \quad (1)$$

with  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ . A panel cointegration test allowing for breaks may be defined very simply as a summary statistic of the cointegration statistics with break computed for the individual units. This problem has been first tackled by Gregory and Hansen (1996), who proposed to compute a no-cointegration statistic  $\theta(t_i^b)$  on the residuals of model (1) for all possible break points  $t_i^b$  and, assuming the rejection region is the left tail (as in the case of the popular *ADF* and *Z* tests), take the minimum. The break point is thus implicitly estimated as  $\hat{t}_i^b = \arg \min(\theta(t_i^b))$ . Another option, favoured by Westerlund and Edgerton (2008) and which we will follow, is to estimate the breakpoint on the basis of a least square criterion (which for stationary variables is consistent even under multiple breaks):  $\hat{t}_i^b = \arg \min(\sum_{t=1}^T \epsilon_{it}^2(t_i^b))$ . For each unit the set of residuals  $\{\epsilon_{it}(\hat{t}_i^b)\}_{t=1}^T$  is then by definition optimal in a least square sense, and can be used to compute a no-cointegration test with break. In fact, let us consider a first order autoregressive equation for the cointegrating residuals:  $\epsilon_{it} = \rho_i \epsilon_{it-1} + \nu_{it}$ . The hypothesis of no cointegration is equivalent to  $H_0: \rho_i = 1$ , and that of no panel cointegration as the same hypothesis for mean or median of this individual statistics, with the latter reflecting more closely the usual de-

inition of the panel null hypothesis. A bootstrap testing algorithm may then proceed as follows:

1. Estimate the breaking cointegrating model (1) on the dataset  $\{y_{it}, x_{it}\}$ ,  $i = 1, \dots, N$ , obtaining for each unit estimates of the coefficients  $(\hat{\mu}_{ji}, \hat{\beta}_{ji}, j = 0, 1)$  and of the optimal cointegrating residuals  $\{\hat{\epsilon}_{it}\}$ ;
2. Compute the  $N$  individual no cointegration statistics  $\hat{\theta}_i$ , on the basis of the the optimal cointegrating residuals  $\{\hat{\epsilon}_{it}(t_i^b)\}$ ;
3. Compute the desired summary statistics (e.g.  $\hat{\theta}_{mean} = N^{-1} \sum_{i=1}^N \hat{\theta}_i$ );
4. Compute  $\hat{\nu}_{it} = \hat{\epsilon}_{it} - \hat{\rho}_i \hat{\epsilon}_{it-1}$ , where  $\{\hat{\epsilon}_{it}\}$  are the estimated cointegrating residuals and  $\hat{\rho}_i$  is a consisten estimate (e.g., OLS) estimate of  $\rho_i$ ;
5. Resample the series  $\{\hat{\nu}_{it}\}$  via the stationary bootstrap, obtaining  $\{\nu_{it}^*\}$ ;
6. Cumulate  $\{\nu_{it}^*\}$  obtaining pseudoresiduals  $\{\epsilon_{it}^*\}$  obeying the null hypothesis of no cointegration;
7. For each unit  $i$  build the pseudodata under the null hypothesis of no cointegration and break in  $\hat{t}_i^b$ :

$$y_{it}^* = \begin{cases} \hat{\mu}_{0i} + \hat{\beta}_0 x_{it} + \epsilon_{it}^*, & t \leq \hat{t}_i^b \\ \hat{\mu}_{1i} + \hat{\beta}_1 x_{it} + \epsilon_{it}^*, & t > \hat{t}_i^b \end{cases} \quad (2)$$

8. Using the datasets  $\{y_{it}^*, x_{it}\}$  estimate the breaking cointegrating model (1) with the re-estimated breakpoints  $\hat{t}_i^{b*}$  given by (??), obtaining sets of estimates of the optimal cointegrating residuals  $\{\hat{\epsilon}_{it}^*\}$ ;
9. Compute the individual bootstrap no cointegration statistics  $\theta_i^*$ ;
10. Compute the bootstrap Group statistics  $\theta_h^*$  ( $h = mean, median$ );
11. Repeat 5-11  $B$  times;
12. Compute the bootstrap significance level of the statistics:  $p(\theta)^* = prop(\theta_h^* < \theta_h)$ .

We evaluated the properties of this test by simulation, finding that it can deliver considerable power gains over pure time series tests, while ensuring good size control. Full details are reported in Di Iorio and Fachin (2010b)

## References

- Banerjee, A., Carrion-i-Silvestre J.L. (2006) Cointegration in Panel Data With Breaks and Cross-Section Dependence, *Working Paper Series* n. 591, European Central Bank. (Revised December, 2006).

- Di Iorio, F., Fachin, S. (2010a) A Panel Cointegration study of the long-run relationship between Savings and Investments in the OECD economies, 1970-2007, *MRPA Working Paper* 25873
- Di Iorio, F., Fachin, S. (2010b) Savings and Investments in the OECD, 1970-2007: a Panel Cointegration test with breaks. *MRPA Working Paper* 3139.
- Gregory, A.W., Hansen, B.E. (1996) Residual-based tests for cointegration in models with regime shifts , *Journal of Econometrics*, 70, 99-126.
- Gutierrez, L. (2010) Simple Tests for Cointegration in Panels with Structural Breaks, *Applied Economics Letters*, 17, 197-200.
- Westerlund, J. (2006a) Testing for panel cointegration with a level break, *Economics Letters*, 91, 27-33.
- Westerlund, J. (2006b) Testing for Panel Cointegration with Multiple Structural Breaks, *Oxford Bulletin of Economics and Statistics*, 68, 101-132.
- Westerlund, J., Edgerton, D. (2008) A Simple Test for Cointegration in Dependent Panels with Structural Breaks, *Oxford Bulletin of Economics and Statistics*, 70, 665-703.

## **Best Estimate e Prediction Error della Riserva Sinistri R.C.Auto in presenza dell'Indennizzo Diretto**

*Paola Fersini<sup>1</sup>, Giuseppe Melisi<sup>2</sup>, Valerio Scacco<sup>2</sup>*

<sup>1</sup>Dipartimento di Scienze Economiche e Aziendali

LUISS Guido Carli

Viale Romania, 32 - 00197 Roma (Italia)

[pfersini@luiss.it](mailto:pfersini@luiss.it)

<sup>2</sup>Dipartimento di Scienze Statistiche

Fac. Ingegneria dell'Informazione, Informatica e Statistica

"SAPIENZA" Università di Roma

Piazzale Aldo Moro 5, 00185 Roma

[Giuseppe.Melisi@uniroma1.it](mailto:Giuseppe.Melisi@uniroma1.it), [Valerio.Scacco@uniroma1.it](mailto:Valerio.Scacco@uniroma1.it)

### ABSTRACT

L'introduzione dell'Indennizzo Diretto nel ramo R.C.A. ha comportato negli ultimi anni una maggiore complessità nell'utilizzo delle metodologie statistico - attuariali per la stima della riserva sinistri, sia per la eterogeneità dei dati disponibili per le valutazioni (gestione dei sinistri ante e post 2007), sia e soprattutto per le diverse dinamiche sottostanti i fattori di rischio che caratterizzano e determinano il costo aziendale complessivo delle Compagnie di assicurazioni, in relazione alla gestione sinistri.

In particolare, il D.P.R. n. 254 del 18 luglio 2006, e successivamente la Convenzione tra Assicuratori per il Risarcimento Diretto (CARD) che ha recepito le norme e stabilito gli accordi, disciplinano i limiti e le modalità di applicazione del risarcimento diretto, prevedendo una nuova struttura del costo dei sinistri, basata sul meccanismo dei Forfait gestionali e debitori. Pertanto il nuovo costo aziendale risulta composto da quattro componenti (sinistri No Card, Sinistri Card, Forfait gestionali e forfait debitori), caratterizzate da evoluzioni aleatorie differenti.

Le quattro tipologie di sinistri elencate presentano, infatti, profili di sviluppo diversi: basti pensare, ad esempio, alla velocità di liquidazione dei sinistri Card, presumibilmente superiore a quella dei sinistri No Card. Tuttavia, a meno che non si effettui una suddivisione a ritroso delle serie storiche dei pagamenti ed un'analisi separata delle stesse, potrebbe verificarsi la situazione in cui in un triangolo run-off compaiano importi relativi a generazioni antecedenti all'entrata in vigore dell'indennizzo diretto, e comprensivi quindi di sinistri di tipologia diversa, che pertanto rischiano di intaccare l'attendibilità di una stima, che si basi appunto su quel triangolo.

In questo lavoro si rende conto dei primi risultati di una ricerca che gli a. stanno svolgendo, al fine di individuare una metodologia che, utilizzando come dato di input le differenti ipotesi che descrivono l'evoluzione delle quattro gestioni, sia capace di quantificare la riserva sinistri separatamente tenendo conto delle caratteristiche di ciascuna tipologia di sinistro.

L'idea è quella di applicare gli aspetti attuariali tipici delle assicurazioni sociali e delle assicurazioni vita alla stima della riserva sinistri.

In particolare, l'intenzione è quella di utilizzare l'impalcatura metodologica utilizzata per la stima dei cash flows prodotti da un portafoglio assicurativo vita (valore attuale medio dei futuri utili e/o perdite generati dal portafoglio assicurativo fino alla sua completa estinzione) attraverso un approccio operativo basato su 'models point' (operando quindi su un numero ridotto di sinistri 'campione', omogenei per caratteristiche, anziché analizzarli analiticamente) e che si rifaccia ai processi evolutivi tipici delle assicurazioni vita/sociali, con stima delle assunzioni relative alla 'movimentazione' del numero dei sinistri (probabilità di eliminazioni, di riapertura, di passaggio tra gestioni differenti) e dei costi medi distinti per ciascuna variabile discriminante individuata (per esempio per zona territoriale, per tipologia di gestione, per tipo di veicolo...).

Attraverso un'applicazione numerica è stato possibile effettuare un confronto dei risultati ottenuti con quelli derivanti dall'applicazione dei tradizionali metodi statistici sia deterministici che stocastici.

Inoltre si accenna al problema della variabilità della riserva sinistri nel suo complesso e si prospetta una soluzione nella determinazione della stessa.

## BIBLIOGRAFIA

- [1] ANIA, *Convenzione di Indennizzo Diretto*, [www.cid-ania.it](http://www.cid-ania.it), 1978.
- [2] ANIA, *Statistica Annuale R.C.AUTO*, 2007 e 2008.
- [3] Association of French Insurers, *Indemnisation directe de l'assuré et de Recours entre Sociétés d'assurance Automobile*, FFSA, [www.ffsa.fr](http://www.ffsa.fr), 1965.
- [4] Autorità Garante per la Concorrenza e il Mercato, *Indagine Conoscitiva sul Settore Assicurativo Autoveicoli, Decisione n° 11891*, 17 Aprile, ([www.agcm.it](http://www.agcm.it)), 2003.
- [5] Autorità per la Concorrenza e il Mercato, *Disciplina del Risarcimento Diretto dei Danni Derivanti dalla Circolazione Stradale, Raccomandazione n° AS324 del 1 Febbraio dal Presidente del Consiglio dei Ministri*, 2006.
- [6] DABONI L., *Lezioni di tecnica attuariale delle assicurazioni contro i danni*, Edizioni Lint, Trieste, 1989.
- [7] ENGLAND P. D., and VERRALL R. J., Paper *Analytic and bootstrap estimates of prediction errors in claims reserving*, Insurance: Mathematics and Economics vol. 25, 1999.
- [8] ENGLAND P.D. and VERRALL R. J., Paper *Stochastic claim reserving in general insurance*, Presented to the Institute of Actuaries, [www.actuaries.org.uk](http://www.actuaries.org.uk), 2002.
- [9] ENGLAND P.D., VERRALL R.J. *Predictive distributions of outstanding liabilities in general insurance*, Annals of Actuarial Science, vol. 1, 2006.
- [10] European Council, *Insurance against civil liability in respect of the use of motor vehicles Directive 72/166/EC*, 1972.
- [11] Faculty of Actuaries in Scotland, *Claims Reserving Manual*, vol. 1, London: Institute of Actuaries, 1997.
- [12] GALLI G., SAVINO C., *Direct reimbursement schemes in compulsory motor liability insurance*, Aprile 2006;
- [13] GLEN BARNETT and BEN ZEHNWIRTH, *Best Estimates for Reserves*, 1999.

- [14] Groupe Consultatif Actuariel Européen, Interim Report *Valuation of Best Estimate under Solvency II for Non-life Insurance*, [www.actuaries.org](http://www.actuaries.org), 2008.
- [15] IAA - Risk Margin Working Group, *Measurement of liabilities for insurance contracts: Current Estimates and Risk Margins*, April 2009.
- [16] Insurer Solvency Assessment Working Party, *A Global Framework for Insurer Solvency*, International Actuarial Association, 2004.
- [17] Insurer Solvency Assessment Working Party of the International Actuarial Association, *A Global Framework for Insurer Solvency Assessment*; I.A.A. International Actuarial Association, 2004.
- [18] ISVAP, *Reserve Requirements and Capital Requirements in Non-Life Insurance. An analysis of the Italian MTPL insurance market by stochastic claims reserving models*, Roma, October 2006.
- [19] Klugman, S. – Panjer, H. – Willmot, G. , *Loss Models. From Data to Decisions*, John Wiley & Sons, INC., 2008.
- [20] MACK T., *Which Stochastic Model is Underlying the Chain-Ladder Model?*, Insurance: Mathematics and Economics, vol.15, 1994.
- [21] MACK T., ASTIN Bulletin, 23, *Distribution-free calculation of the standard error of chain-ladder reserve estimates*, Casualty Actuarial Society, [www.casact.org](http://www.casact.org), 1994.
- [22] MACK T., Paper *Measuring the Variability of Chain-Ladder Reserve Estimates*, Casualty Actuarial Society, Spring Forum, 1994.
- [23] RENSHAW, A.E., VERRALL R. J., Paper *A stochastic model underlying the chain ladder technique*, British Actuarial Journal, vol. 4, 1998.
- [24] VERRALL R. J., *Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion*, Casualty Actuarial Society, 2002.

## **NORMATIVA E REGOLAMENTI**

Decreto del Ministro dello Sviluppo Economico 11 Dicembre 2009,

Decreto del Ministero dello Sviluppo Economico 1 aprile 2008, n. 86

Relazione del Comitato tecnico ministeriale di cui all'art. 13 D.P. R. n. 254/2006, con D.M. 19 dicembre 2006

Decreto del Presidente della Repubblica 18 luglio 2006, n. 254

## **Gli effetti della dipendenza dei rischi sul pricing di una copertura congiunta incendio e furto ed r.c.auto**

*Salvatore Forte, Matteo Ialenti, Marco Pirra*

Nell'ambito dei prodotti dei rami danni è frequente, nel mercato assicurativo italiano, la vendita congiunta di prodotti, quali ad esempio la responsabilità civile auto e l'incendio e furto. In tale casistica, a parità di contraente, la Compagnia assicura più di un rischio e per questo motivo ci potrebbero essere effetti di dipendenza tra i rischi assicurati.

L'obiettivo del lavoro è la costruzione di un modello di danno aggregato finalizzato al pricing della garanzie singole e della garanzia congiunta.

Per lo studio delle dipendenze si è fatto ricorso all'utilizzo delle funzioni copula.

Il modello attuariale permette, oltre al pricing delle coperture, anche una valutazione prospettica dell'utile tecnico e una valutazione del nuovo requisito patrimoniale per il non-life premium risk in ottica Solvency II.

## **Solvency II: modelli interni per le riserve sinistri, tecniche di Re-reserving e Backtesting**

*Salvatore Forte, Fabio Grasso, Matteo Ialenti, Marco Pirra*

Le riserve sinistri rappresentano, quasi sempre, la posta più importante dello stato patrimoniale passivo per una Compagnia di Assicurazioni contro i danni. Per questo motivo una adeguata stima delle riserve sinistri da parte dell' Attuario, per una Compagnia operante in più rami ministeriali danni, rappresenta una delle attività più delicate. Tenendo presente i principi contabili nazionali e il Regolamento ISVAP n. 16, la stima prevede una valutazione a "costo ultimo", ovvero a valore nominale, in un'ottica prudenziale.

Le nuove regole europee in tema di vigilanza (Solvency II) impongono una valutazione delle riserve sinistri a valori di mercato (full balance sheet approach). Tale valutazione corrisponde alla somma di due componenti: la best estimate e il margine per il rischio.

Quest'ultima componente deve essere valutata mediante l'applicazione del principio del costo del capitale, che necessita, a sua volta, di una misura di rischio multi periodale.

Nella letteratura di settore sono state proposte differenti approssimazioni data la complessità di tale oggetto matematico.

Nel lavoro si propone una metodologia per valutare il margine per il rischio, la best estimate e il nuovo requisito patrimoniale mediante l'applicazione di un modello interno in chiave stocastica.

Al fine di testare la metodologia proposta sono state considerate tre Compagnie che operano nel mercato della responsabilità civile generale e che si differenziano tra di loro nella dimensione del portafoglio sinistri.

Nell'ambito del modello interno, sono state proposte differenti metodologie stocastiche; al fine di scegliere un solo metodo per la valutazione delle riserve sinistri è stato altresì implementato un processo di backtesting.

# SPECTRAL ANALYSIS OF SQUARE MATRIX POLYNOMIALS BY LOCAL RANK FACTORIZATION

MASSIMO FRANCHI AND PAOLO PARUOLO

**Abstract.** A method of constructing the Smith form as well as associated equivalence transformations for square matrix polynomials  $A(\lambda)$  is presented; the method employs a finite sequence of ‘local rank factorizations’ and it is shown to provide a complete and explicit description of Jordan pairs. When  $A(\lambda) = A - \lambda I$ , the analysis delivers the Jordan form of  $A$  and a Jordan basis.

**Key words.** Matrix polynomials; spectral theory; canonical forms; Smith form; Jordan form; Jordan chains; equivalence of matrix polynomials; similarity of matrices

**AMS subject classifications.** 15A09, 15A21, 47A46, 47A56

**1. Introduction.** It is well known, see Gantmacher [6], that every matrix polynomial has a diagonal canonical form, i.e.  $A(\lambda) = U(\lambda)\Lambda(\lambda)V(\lambda)$ , where  $\Lambda(\lambda)$  is called the Smith form of  $A(\lambda)$ . In this paper we present a method of constructing  $\Lambda(\lambda)$ ,  $U(\lambda)$  and  $V(\lambda)$  for square matrix polynomials. The method builds on the ‘local rank factorization’ (LRF) introduced in [4] and it is shown to provide a complete and explicit description of Jordan pairs. When  $A(\lambda) = A - \lambda I$ , the analysis delivers the Jordan form of  $A$  and a Jordan basis, i.e. a system of generalized eigenvectors that transforms the matrix into its canonical form.

These results extend the study of the spectral theory of matrix polynomials, see Gohberg, Lancaster and Rodman [7], which has important applications in matrix theory [6, 14, 16], in the study of systems of differential and difference equations [7, 8, 13], in system theory [11, 10, 2], in operator theory [12, 15] as well as in times-series econometrics [9, 3, 5]. The same tools are also employed in numerical algorithms, such as the ones in [1, 17, 18] for calculating the coefficients of the principal part of the inverse.

**2. Conclusion.** One can view the results of the paper as an alternative proof of the possibility of reducing a matrix polynomial to its Smith form; however, since the canonical form is found by non elementary transformations, a different notion of equivalence of matrix polynomials must be introduced.

## REFERENCES

- [1] K. AVRACHENKOV, M. HAVIV, AND P. HOWLETT, *Inversion of analytic matrix functions that are singular at the origin*, SIAM Journal of Matrix Analysis and Applications, 22 (2001), pp. 1175–89.
- [2] H. BART, I. GOHBERG, M. KAASHOEK, AND A. RAN, *Factorization of Matrix and Operator Functions: The State Space Method*, Birkhäuser, Basel-Boston-Berlin, 2008.
- [3] M. FRANCHI, *A representation theory for polynomial cofractionality in vector autoregressive models*, Econometric Theory, 26 (2010), pp. 1201–1217.
- [4] M. FRANCHI AND P. PARUOLO, *Inversion of regular analytic matrix functions: local Smith form and subspace duality*, wp, 2010.
- [5] ———, *A characterization of vector autoregressive processes with common cyclical features*, Journal of Econometrics, (in press).
- [6] F.R. GANTMACHER, *The Theory of Matrices*, vol. I, AMS Chelsea Publishing, 1959.
- [7] I. GOHBERG, P. LANCASTER, AND L. RODMAN, *Matrix polynomials*, Academic Press, New York, 1982.
- [8] ———, *Invariant subspaces of matrices with applications*, Wiley, New York, 1986.
- [9] N. HALDRUP AND M. SALMON, *Representation of  $I(2)$  cointegrated systems using the Smith-McMillan form*, Journal of Econometrics, 44 (1998), pp. 303–325.
- [10] P. HOWLETT, *Input retrieval in finite dimensional linear systems*, Journal of Australian Mathematical Society (Series B), 23 (1982), pp. 357–82.
- [11] T. KAILATH, *Linear systems*, Prentice Hall, Englewood Cliffs, N.J., 1980.
- [12] M. KELDYSH, *On eigenvalues and eigenfunctions of certain classes of non self-adjoint operators*, Dokl. Akad. Nauk SSSR (Russian), 77 (1951), pp. 11–14, English transl. in [15].
- [13] P. LANCASTER AND M. TISMENETSKY, *The theory of matrices*, Academic Press, 1985.
- [14] C. LANGENHOP, *The Laurent expansion for a nearly singular matrix*, Linear Algebra and its Applications, 4 (1971), pp. 329–40.
- [15] A. MARKUS, *Introduction to the spectral theory of polynomial operator pencils*, *Translations of Mathematical Monographs*, vol. 71, Americal Mathematical Society, Providence, 1988.
- [16] P. SCHWEITZER AND G. STEWART, *The Laurent expansion of pencils that are singular at the origin*, Linear Algebra and its Applications, 183 (1993), pp. 237–54.
- [17] J. WILKENING, *An algorithm for computing Jordan chains and inverting analytic matrix functions*, Linear Algebra and its Applications, 427 (2007), pp. 6–25.
- [18] J. WILKENING AND J. YU, *A local construction of the Smith normal form of a matrix polynomial*, Journal of Symbolic Computation, (in press).

# Computing Graph Spanners in Small Memory: Fault-Tolerance and Streaming <sup>★</sup>

Giorgio Ausiello<sup>1</sup>, Paolo G. Franciosa<sup>2</sup>, Giuseppe F. Italiano<sup>3</sup>, and Andrea Ribichini<sup>1</sup>

<sup>1</sup> Dipartimento di Informatica e Sistemistica, Sapienza Università di Roma, via Ariosto 25, I-00185 Roma, Italy. {ausiello,ribichini}@dis.uniroma1.it

<sup>2</sup> Dipartimento di Scienze Statistiche, Sapienza Università di Roma, piazzale Aldo Moro 5, I-00185 Roma, Italy. paolo.franciosa@uniroma1.it

<sup>3</sup> Dipartimento di Informatica, Sistemi e Produzione, Università di Roma “Tor Vergata”, via del Politecnico 1, 00133 Roma, Italy. italiano@disp.uniroma2.it

A spanner of an undirected graph is a subgraph on the same vertex set which preserves approximate distances between all pairs of vertices. More formally, given  $\alpha \geq 1$  and  $\beta \geq 0$ , an  $(\alpha, \beta)$ -spanner of a graph  $G(V, E)$  is a subgraph  $S(V, E')$  of  $G$  such that  $d_S(x, y) \leq \alpha \cdot d_G(x, y) + \beta$  for each pair of vertices  $x, y \in V$ , where  $d_G(x, y)$  represents the distance between  $x$  and  $y$  in graph  $G$ .

The interest towards graph spanners arises mainly from their sparsity: approximating distances by a multiplicative factor, or even by simply an additive term, allows one to reduce substantially the space requirements. Indeed, it is known (see [7]) that any graph with  $n$  vertices contains a  $(k, k - 1)$ -spanner with  $O(n^{1+1/k})$  edges.

Another interesting notion introduced in the literature is the concept of fault-tolerant spanner: given an integer  $f \geq 1$ , an  $f$ -vertex fault-tolerant spanner of a graph is a subgraph that preserves distances, within the same distortion, under the failure of any set of at most  $f$  vertices. More formally, an  $f$ -vertex fault-tolerant  $(\alpha, \beta)$ -spanner of  $G(V, E)$  is a graph  $S(V, E')$ , with  $E' \subseteq E$ , such that for each subset of vertices  $F \subseteq V$ , with  $|F| \leq f$ , and for each pair of vertices  $x, y \in V \setminus F$  we have  $d_{S \setminus F}(x, y) \leq \alpha \cdot d_{G \setminus F}(x, y) + \beta$ , where by  $G \setminus F$  we mean the subgraph of  $G(V, E)$  induced by vertex set  $V \setminus F$ . An  $f$ -edge fault-tolerant spanner is defined analogously, under the failure of at most  $f$  edges.

Due to their sparsity, graph spanners appear to be useful especially when dealing with very large and dense graphs. In this scenario, it is quite natural to assume that the input graph must be read from an external device and that it cannot be stored in internal memory.

*State of the art:* Several algorithms have been proposed in the literature for computing graph spanners in the data streaming model [2, 3, 5, 6]. In this model, input data are scanned as a sequence of values in arbitrary order, and the storage available to the algorithm is smaller than the input size. Thus, the algorithm is not allowed to access randomly the input data. Given a graph with  $n$  vertices, the spanners computed by all these algorithms have  $\Omega(n^{1+1/k})$  edges, where  $2k + 1$  is the desired multiplicative stretch factor. Slightly better stretch/size trade-offs can be obtained in case of small stretch spanners: the  $(k, k - 1)$ -spanners in [2], for  $k \in \{2, 3\}$ , have  $O(n^{1+1/k})$  edges. An extensive study of the experimental behavior of the above algorithms is in [1].

While fault-tolerant spanners for geometric graphs have been studied in the last decade, they have only recently been introduced for general graphs by Chechik *et al.* [4]. Their construction builds an  $f$ -vertex fault-tolerant  $(2k - 1, 0)$ -spanner containing, with high probability,  $O(f^2 k^{f+1} \cdot n^{1+1/k} \log^{1-1/k} n)$  edges. The total time required to compute the spanner is  $O(f^2 k^{f+1} \cdot n^{3+1/k} \log^{1-1/k} n)$ .

*Our results:* The contribution of this paper is twofold: we introduce practical algorithms for computing fault-tolerant small stretch spanners for general graphs, which improve substantially on previously known bounds [4] and propose, for the first time, streaming algorithms for computing spanners in very small (i.e., even as small as  $\Theta(B)$ ) internal memory. We assume that the input is read in a single pass on a data stream, and the computation can exploit only a very small internal memory plus an external memory device from/to which data are transferred in blocks of size  $B$ , like in the classical EM

<sup>★</sup> The full paper appeared in Discrete Mathematics, Algorithms and Applications, Dec 2010.

model [8]. We measure the performance of our algorithms in terms of the number of block transfers. Since we are interested in dealing with very large graphs, and some of today’s applications are required to run even on small computing devices (such as mobile devices and smart phones), it seems crucial to use a limited amount of internal memory, and to store most of the information needed on slow access external memory devices.

Let  $G = (V, E)$  be an undirected unweighted graph, with  $m$  edges and  $n$  vertices. Our streaming algorithms compute (3,2)- and (2,1)-spanners of  $G$  having asymptotically optimal size: the (3,2)-spanner contains  $O(n^{4/3})$  edges, while the (2,1)-spanner contains  $O(n^{3/2})$  edges. Both algorithms work with an internal memory of size  $s = \Theta(\max(\log n, B))$  plus an external memory proportional to the size of the spanner, and require a single pass on the input stream, in which graph edges may appear in any arbitrary order. The I/O complexity of our algorithms is  $\tilde{O}(m/B)$  in the worst case (throughout the paper, we use the notation  $\tilde{O}(f(n)) = O(f(n)\text{polylog}n)$ ), and hence it is optimal up to a polylogarithmic factor. Our algorithms do not require advance knowledge of the number  $n$  of vertices or the number  $m$  of edges in the graph. While processing the input stream, the algorithms build an optimal spanner on the portion of the graph examined so far: in all the complexity bounds,  $n$  and  $m$  refer to the number of vertices and edges currently known to the algorithm. We remark that data stored on external memory are accessed only by sequential scans and sorting, and thus the algorithms are expected to be very efficient in practice.

We also provide algorithms for computing  $f$ -vertex fault-tolerant (3,2)- and (2,1)-spanners of general graphs, which can be built in the same computational model described above. The (3,2)-spanner contains  $O(f^{4/3}n^{4/3})$  edges and can be computed with  $\tilde{O}(f^2m/B)$  I/O’s in the worst case. The (2,1)-spanner contains  $O(fn^{3/2})$  edges and can be computed with  $\tilde{O}(fm/B)$  I/O’s in the worst case. In the classical RAM model, the computations of our  $f$ -vertex fault-tolerant (3,2)- and (2,1)-spanners require respectively only time  $\tilde{O}(f^2m)$  and  $\tilde{O}(fm)$ . The corresponding fault-tolerant spanners of Chechik *et al.* [4] achieve the following bounds: their fault-tolerant (5,0)-spanner contains  $O(f^23^{f+1} \cdot n^{4/3} \log^{2/3} n)$  edges with high probability and can be computed in time  $\tilde{O}(f^23^{f+1} \cdot n^{10/3})$ , while their fault-tolerant (3,0)-spanner contains  $O(f^22^{f+1} \cdot n^{3/2} \log^{1/2} n)$  edges with high probability and can be computed in time  $\tilde{O}(f^22^{f+1} \cdot n^{7/2})$ . Thus, our algorithms improve significantly the bounds in [4] in the case of small stretch spanners.

*Conclusions* Besides improving any of our bounds, there are few other issues that seem worth further investigation. First, can we extend our approach to larger stretch spanners? Second, a thorough experimental study may shed some light on practical approaches to compute fault-tolerant spanners.

## References

1. G. Ausiello, C. Demetrescu, P. G. Franciosa, G. F. Italiano, and A. Ribichini. Graph spanners in the streaming model: An experimental study. *Algorithmica*, 55(2):346–374, 2009.
2. G. Ausiello, P. G. Franciosa, and G. F. Italiano. Small stretch  $(\alpha, \beta)$ -spanners in the streaming model. *Theoretical Computer Science*, 410(36):3406–3413, 2009.
3. S. Baswana. Streaming algorithm for graph spanners - single pass and constant processing time per edge. *Inf. Process. Lett.*, 106(3):110–114, 2008.
4. S. Chechik, M. Langberg, David Peleg, and Liam Roditty. Fault-tolerant spanners for general graphs. In *Proc. of 41st Annual ACM Symposium on Theory of Computing (STOC’09)*, pages 435–444, 2009.
5. M. Elkin. Streaming and fully dynamic centralized algorithms for constructing and maintaining sparse spanners. In *Proc. of the 34th International Colloquium on Automata, Languages and Programming (ICALP’07)*, volume 4596 of *LNCS*, pages 716–727. Springer, 2007.
6. M. Elkin and J. Zhang. Efficient algorithms for constructing  $(1 + \varepsilon, \beta)$ -spanners in the distributed and streaming models. *Distributed Computing*, 18(5):375–385, 2006.
7. D. Peleg and A. Schäffer. Graph spanners. *Journal of Graph Theory*, 13:99–116, 1989.
8. J. S. Vitter. External memory algorithms and data structures: dealing with massive data. *ACM Computing Surveys*, 33(2):209–271, 2001.

## Strategie di ricerca con analisi automatica dei dati testuali

Luca Giuliano

In un recente convegno promosso dalla Fondazione Telecom Italia al Politecnico di Torino (2060: *con quali fonti si farà la storia del nostro presente?* 8-9 aprile 2010) storici, sociologi, archivisti, filosofi e informatici si sono interrogati sui problemi connessi con quanto del nostro presente è già consegnato al futuro in formato digitale per il solo fatto di essere nato nella “società in rete” (comunicazioni interpersonali, transazioni economiche e finanziarie, pubblicazioni e attività di ricerca scientifica, cronaca quotidiana e dibattito politico, solo per ricordarne una piccola parte). Ma a questo si aggiunge il crescente impegno con cui archivi e biblioteche stanno digitalizzando in modo retrospettivo testi e documenti analogici che fanno parte del loro patrimonio e che per la prima volta verranno messi a disposizione di una consultazione di massa.

Le procedure informatiche e la digitalizzazione delle conoscenze rappresentano la “nuova mnemotecnica” del nostro tempo: un sussidio artificiale al servizio dell’intelligenza. L’immenso contributo delle nuove tecnologie digitali nella gestione di grandi masse di informazioni e dati ha interessato tutte le discipline empirico-formali che puntano a un certo grado di generalizzazione dei risultati. Le conseguenze della digitalizzazione sono state decisive per il modo di concepire i “documenti” di cui le scienze sociali fanno largo uso e che costituiscono le basi empiriche di controllo delle ipotesi di ricerca. I documenti di ricerca, se non sono preparati nel momento della loro rilevazione in una forma strutturata e classificabile a priori, come accade in un test o in un questionario i cui risultati sono trasferiti immediatamente in codici alfanumerici, sono irriducibilmente ancorati alla loro forma fisica originaria: la lettera, l’articolo di giornale, il manifesto, il diario, la fotografia ecc. I documenti analogici, prodotti per qualsiasi motivo nella vita sociale e solo in un secondo momento portati all’attenzione dello studioso sulla base di specifici interessi e scelte di ricerca, non possono essere sottoposti alla generalizzazione empirica se non con una immensa fatica e con un enorme impiego di tempo. Thomas e Znaniecki nel 1920 portarono a termine il loro lavoro *The Polish Peasant in Europe and America* analizzando in sei anni almeno mille lettere e ottomila documenti vari tratti da giornali dell’epoca, ma nessuno sa esattamente quanti fossero, quali furono scartati e perché. Questi problemi furono messi in luce molto tempo dopo la pubblicazione, in un convegno del 1938 durante il quale si seppe che gran parte della documentazione originale era andata distrutta. Una base di dati fondata sull’analisi dei documenti naturali, tipicamente d’archivio, ha un carattere difficilmente ispezionabile, se non riproducendo il percorso, spesso singolare e intuitivo, dello studioso che lo ha compiuto.

La digitalizzazione dei documenti, in particolare dei testi, può cambiare radicalmente questa situazione. I testi possono essere condivisi tra diversi ricercatori. L’analisi del contenuto può essere condotta con software CAQDAS (Computer Assisted Qualitative Data Analysis Softwares) appositamente sviluppati per seguire il percorso di concettualizzazione, di operativizzazione e di classificazione. Testi non strutturati, come le interviste, i focus group, i forum in Internet, le email o i messaggi in Facebook, possono essere classificati automaticamente ed essere restituiti all’esame critico della comunità scientifica dotati di tutte le meta-informazioni che permettono la ricostruzione e la trasparenza delle scelte metodologiche compiute dai ricercatori.

L’associazione e l’incontro tra l’analisi qualitativa assistita dal computer e l’analisi statistica dei dati testuali rendono operativa l’integrazione tra comprendere e spiegare: tra il lavoro di interpretazione che si focalizza sul testo, sulla intenzionalità dell’autore, sull’universo che in esso si esprime e sulla sua singolare costruzione di senso con il lavoro di spiegazione che opera un distacco del testo dal suo autore, che decontestualizza le parole per portarle a sostegno di ipotesi di ricerca, sottoporle a confronti, misure, formalizzazioni che ne permettono la generalizzazione.

Questa prospettiva di ricerca e analisi permette l'incontro tra scienze che hanno interesse a mettere in evidenza le proprietà singolari e uniche degli oggetti e scienze che hanno interesse a selezionare le proprietà che costituiscono classi di oggetti. Il ricercatore si muove così dal testo preso nella sua completezza per risalire alle parole che costituiscono il testo. Questo "doppio movimento" dalla qualità alla quantità e dalla quantità alla qualità diventa essenziale e irrinunciabile quando il ricercatore ha a disposizione basi testuali e corpora costituiti da centinaia di migliaia o da milioni di occorrenze. Allora, nella impossibilità di leggere il testo per interpretarne il senso egli mette in atto strategie lessicometriche e quantitative per individuare parole-chiave, segmenti ripetuti, unità semantiche che presentano qualche interesse rilevante, qualche peculiarità di presenza all'interno del corpus. Poi, successivamente, assume un punto di vista ermeneutico quando seleziona le concordanze o individua nel corpus parti di testo che presentano un interesse particolare in base alla presenza delle parole chiave individuate.

L'analisi statistica dei dati testuali è in grado di offrire una base empirica sistematica ad osservazioni compiute con l'analisi dei dispositivi retorici della comunicazione, integrando metodologie qualitative e quantitative che appaiono di importanza fondamentale in queste strategie di analisi che si collocano in un ambito socio-linguistico.

Il software dedicato al riconoscimento automatico dei dati espressi in linguaggio naturale per mezzo di dizionari elettronici, taggers sintattici e grammaticali, motori di lemmatizzazione, risolve a priori le principali ambiguità legate alla lingua e permette l'esplorazione dei testi per l'estrazione dell'informazione essenziale secondo gli obiettivi del ricercatore. In parte, le tecniche adottate si riconducono all'ambito più recente del text mining che - a partire dalla ricerca di classi di parole e di concetti espressi mediante categorie semantiche e trovati attraverso la formulazione di espressioni regolari in grado di catturare sequenze ripetute, co-occorrenze, strutture a elementi variabili o altri gruppi lessicali - permette di strutturare informazioni libere e sparse nei testi all'interno di matrici di dati tradizionali e strutturati. Queste ultime sono poi analizzabili con tecniche multidimensionali che elaborano una grande massa di dati ad alta variabilità.

Le linee metodologiche di ricerca fin qui indicate sono sviluppate in due progetti finanziati dei quali lo scrivente è responsabile scientifico:

- "La rappresentazione sociale della Cina nella stampa quotidiana europea in relazione con le dinamiche economiche e migratorie: mutamenti e confronti tra Italia, Spagna, Regno Unito e Francia attraverso le tecniche di analisi automatica dei dati testuali." (quota MIUR 2007 - assegnazioni di Ateneo Federato della Scienza e della Tecnologia; altri comp. del Dipartimento Scienze statistiche: V. Egidi, F. Deriu).

- "Il linguaggio della leadership politica tra la prima e la seconda Repubblica. Analisi statistica e linguistica dei discorsi parlamentari" (quota MIUR 2009 - assegnazioni di Sapienza - Università di Roma; altri comp. del Dipartimento di Scienze statistiche: G. Salinetti).

All'interno del primo progetto sono stati costituiti tre corpora di linguaggio giornalistico per la lingua italiana, inglese, e spagnola. Dalle prime analisi è emerso, in riferimento al tema della valutazione positiva o negativa dell'immagine della Cina, un problema di classificazione automatica delle forme grafiche sotto il profilo della "sentiment analysis" o "opinion mining" che è in corso di definizione comparando liste di forme grafiche secondo le categoria grammaticali di assegnazione. La valutazione di sintesi, già abbastanza complessa nel dover tenere conto del contesto, appare molto sensibile alle variazioni di categoria grammaticale.

Il secondo progetto è nella fase di costruzione del corpus che è parte integrante e basilare del progetto, dal momento che non si dispone attualmente di un corpus di riferimento del linguaggio politico sotto nessun punto di vista. L'ipotesi di lavoro, privilegiando il tipo di discorso (in particolare: informativo, persuasivo, narrativo) rispetto ai contenuti della comunicazione stessa e anche delle forme sintattiche e grammaticali utilizzate, assume che il linguaggio politico

contemporaneo si sia in parte tecnicizzato, assumendo modalità strumentali, al punto che la dimensione ideologica che ne era la base di strutturazione fino agli anni '70 è stata sostituita da una dimensione più economica e pragmatica. All'inizio degli anni '90, in corrispondenza con la crescente complessità del sistema politico, anche in funzione delle scelte imposte dalla nascente Unione Europea, la “politica dei burocrati” ha comportato una diminuzione dell'interesse e dell'impegno, testimoniati anche dall'aumento dell'astensionismo elettorale. Il linguaggio non è più portatore di ideologie e le differenze programmatiche sfumano a favore di un pensiero unico che valorizza i condizionamenti della struttura economica a discapito dal dibattito etico-normativo.

L'analisi delle strategie discorsive messe in atto dai diversi leader politici dovrebbe rendere conto di questi mutamenti e dell'eventuale recupero delle componenti emozionali di questi ultimi anni (1994-2008) in cui la semplificazione del linguaggio punta anche a una semplificazione della politica e alla costruzione del consenso. Alcuni studi recenti hanno messo in evidenza la manipolazione del sistema pronominale in relazione alla distribuzione delle responsabilità politiche o della formazione/consolidamento dell'identità. In altri casi è stata ipotizzata la possibilità di riconoscere il tipo di discorso dai tempi dei verbi (per esempio: il tempo futuro per la progettualità, il tempo passato per la narrazione).

### **Pubblicazioni di riferimento legate ai due progetti di ricerca:**

- Giuliano L. (2010). L'immagine della Cina nella stampa quotidiana dal 2000 al 2008. In L. Giuliano e G. La Rocca, *L'analisi automatica e semi-automatica dei dati testuali. Vol. II – Strategie di ricerca e applicazioni*, Milano: LED, pp. 93-113.
- Giuliano L., La Rocca G. (2010). Validity and reliability of the automatic classification of texts according to the negative-positive criterion. In: *Statistical Analysis of Textual Data*. Sapienza University of Rome, 2010 9-11 June, Milano: LED Ed. Univ. di Lettere Economia e Diritto, vol. 1, pp. 61-72..
- Galli de' Paratesi N., Giuliano L. (2009). Pronoun morphology, modality and semantics of political communication in presidential debate of two Italian political leaders. *Révue internationale de sociologie*, vol. 19 n. 3; pp. 401-432.
- Giuliano L. (2008). Parole e politica nei «faccia a faccia» della campagna elettorale del 2006. In: G. Sensales e M. Bonaiuto (a cura di). *La politica mediatizzata. Forme della comunicazione politica nel confronto elettorale del 2006*. pp. 163-189, Milano: Franco Angeli.
- Bonaiuto M., Giuliano L., Maricchiolo F. (2008). Parole e gesti nell'analisi automatica del testo: il caso dei faccia a faccia tra Berlusconi e Prodi. In: *JADT-2008 - Actes des 9èmes Journées Internationales d'Analyse Statistiques des Données Textuelles*. Lyon, 12-14 mars, Lyon: Presses universitaires de Lyon, vol. 2, pp. 787-798.

## Data Augmentation Approach in Bayesian Modelling of Presence-only Data

F. Divino, N. Golini, G. Jona Lasinio, A. Penttinen

**Keywords :** Bayesian model, Data augmentation, Presence only-data

### Abstract

An important issue in ecological studies is to estimate the potential spatial extent of an ecological niches. The prediction of geographic distribution of species in suitable areas is essential for planning conservation and management strategies, and it may concern animals and plants species studies. Given the presence-absence data for a species, a logistic model represents a natural approach to estimate the species distribution, given the environmental covariates. In ecological studies, however, collecting such presence-absence data could be expensive and/or difficult. Often the only available information is the true presence of the species and the associated environmental covariates in some locations of the study area. Such data are known in literature as presence-only data, see Pearce and Boyce (2006). Following the pseudo-absence approach (Ward et al., 2009), it is possible to adapt standard case-control models to presence-only data. We propose a Bayesian model to estimate logistic linear regressions adapted to presence-only data of rare species. We introduce a random approximation of the correction factor in the adjusted logistic model that allows us to overcome the need to know the prevalence of the species. Under the assumption that the pseudo-absence locations are randomly sampled from the entire study area, we can estimate regression parameters jointly with prevalence through a data augmentation MCMC algorithm (Liu, 2004). Preliminary simulations investigating the behaviour of our model with respect to the spatial structure of the covariates, to the rareness of the species and to the dimension of the pseudo-absences sample have shown encouraging results. References: - Liu S. J. (2004), "Monte Carlo Strategies in Scientific Computing", Springer, New York. - Pearce, J. L. and Boyce, M. S. (2006), "Modelling distribution and abundance with presence-only data", *Journal of Applied Ecology*, 43, 405-412. - Ward, G., Hastie, T., Barry, S., Elith, J. and Leathwick, A. (2009), "Presence-only data and the EM algorithm", *Biometrics*, 65, 554-563.

# Pseudo processi di Fresnel ed equazioni di ordine superiore

Enzo Orsingher      Mirko D'Ovidio

Lo studio dei processi governati da misure segnate ha interessato diversi ricercatori sia per gli strumenti matematici che tale studio impegna che per lo stretto legame con la fisica matematica. Inoltre, molti funzionali di tali processi sono delle variabili aleatorie definite su spazi con misure di probabilità. Tali processi sono governati da equazioni di ordine superiore del tipo del calore

$$\frac{\partial u}{\partial t}(x, t) = \kappa_n \frac{\partial^n u}{\partial x^n}(x, t), \quad x \in \mathbb{R}, \quad t > 0, \quad n \in \mathbb{N}$$

con dato iniziale  $u(x, 0) = \delta(x)$ .

Un'equazione fondamentale della fisica matematica è l'equazione delle vibrazioni delle sbarre

$$\frac{\partial^2 u}{\partial t^2}(x, t) = -\kappa^2 \frac{\partial^4 u}{\partial x^4}(x, t), \quad x \in \mathbb{R}, \quad t > 0$$

che possiede interessanti relazioni con l'equazione fondamentale della meccanica quantistica (equazione di Schrödinger) essendo

$$\left( \frac{\partial^2}{\partial t^2} + \frac{1}{2} \frac{\partial^4}{\partial x^4} \right) = \left( \frac{\partial}{\partial t} - \frac{i}{2} \frac{\partial^2}{\partial x^2} \right) \left( \frac{\partial}{\partial t} + \frac{i}{2} \frac{\partial^2}{\partial x^2} \right)$$

e con l'equazione del calore (a sua volta legata con gli pseudo processi di ordine  $n = 4$ ). Sotto le condizioni iniziali  $u(x, 0) = \delta(x)$  e  $u_t(x, 0) = 0$  la soluzione dell'equazione delle sbarre è data dal nucleo

$$K(x, t) = \frac{1}{\sqrt{2\pi t}} \cos\left(\frac{x^2}{2t} - \frac{\pi}{4}\right), \quad x \in \mathbb{R}, \quad t > 0$$

che riveste un ruolo fondamentale nel presente lavoro. Tale nucleo può essere associato alla misura segnata  $\mu(dx, t) = K(x, t)dx$  governante lo pseudo processo  $X_{sb}(t)$ ,  $t > 0$  per cui è possibile identificare una versione  $n$ -dimensionale di tale misura. Qui diamo il caso bidimensionale

$$\mu(dx_1, t_1; dx_2, t_2) = \frac{dx_1 dx_2}{2\pi \sqrt{t_1(t_2 - t_1)}} \sin\left(\frac{x_1^2}{2t_1} + \frac{(x_2 - x_1)^2}{2(t_2 - t_1)}\right).$$

Si possono studiare diversi problemi al contorno per l'equazione delle sbarre. Consideriamo  $\kappa = 1/2$ :

- sotto la condizione assorbente

$$\begin{cases} u(x, 0) = \delta(x - y), & x, y > 0 \\ u_t(x, 0) = 0 \\ u(0, t) = 0, & t > 0 \end{cases} \quad (0.1)$$

si ottiene

$$u(x, t) = K(x - y, t) - K(x + y, t)$$

- sotto la condizioni elastica

$$\begin{cases} u(x, 0) = \delta(x - y), & x, y > 0 \\ u_t(x, 0) = 0 \\ u_x - \alpha u|_{x=0} = 0, & \alpha > 0, t > 0 \end{cases} \quad (0.2)$$

si ottiene

$$u(x, t) = K(x - y, t) - K(x + y, t) - 2\alpha e^{\alpha(x+y)} \int_{x+y}^{\infty} e^{-\alpha w} K(w, t) dw$$

Si è studiata inoltre la versione frazionaria dell'equazione delle sbarre

$$\frac{\partial^{2\nu} u}{\partial t^{2\nu}}(x, t) = -\frac{1}{2^2} \frac{\partial^4 u}{\partial x^4}(x, t), \quad x \in \mathbb{R}, t > 0$$

sogetta alle condizioni iniziali

$$\begin{cases} u(x, 0) = \delta(x) \\ u_t(x, 0) = 0 \end{cases}$$

la cui soluzione ha trasformata di Fourier che è la Mittag-Leffler generalizzata

$$\hat{u}(\beta, t) = E_{2\nu, 1} \left( -\frac{\beta^4 t^{2\nu}}{2^2} \right).$$

Il nucleo  $K(x, t)$  randomizzato da un tempo Browniano rappresenta la soluzione dell'equazione di ordine superiore

$$\frac{\partial u}{\partial t}(x, t) = -\frac{1}{2^3} \frac{\partial^4 u}{\partial x^4}(x, t), \quad x \in \mathbb{R}, t > 0$$

dove  $u(x, t) = E[K(x, |B(t)|)]$ . Come accade per alcuni funzionali dei pseudo processi anche dal processo  $X_{sb}(t)$ ,  $t > 0$  governato dalla misura segnata  $\mu$  si può arrivare a delle vere variabili aleatorie. Infatti, randomizzando il tempo del nucleo  $K$  rispetto ad un tempo di primo passaggio

$$T_t = \inf\{s \geq 0 : B(s) = t\}$$

( $B$  è un moto Browniano standard) si ottiene la legge

$$u(x, t) = E[K(x, T_t)] = \frac{t}{\sqrt{2\pi}} \frac{t^2 + x^2}{t^4 + x^4}, \quad x \in \mathbb{R}, t > 0$$

Un risultato interessante è dato dalla seguente convoluzione

$$(K * K)(x, t) = \frac{x}{2t} \left[ \sin \frac{x^2}{2t} + J_0 \left( \frac{x^2}{2t} \right) \right], \quad x, t > 0$$

che rappresenta la legge di  $\sqrt{{}_1X_{sb}(t)^2 + {}_2X_{sb}(t)^2}$ ,  $t > 0$  dove  ${}_jX_{sb}$ ,  $j = 1, 2$  sono due processi indipendenti governati dalla misura segnata  $\mu$ .  $J_0$  è la funzione di Bessel del primo tipo

$$J_\nu(x) = \sum_{k \geq 0} (-1)^k \frac{(x/2)^{2k+\nu}}{k! \Gamma(k + \nu + 1)}.$$

## References

- [1] E. Orsingher. Brownian fluctuations in space-time with applications to vibrations of rods. *Stochastic Processes and Their Applications* 23 (1986) 221 - 234
- [2] W.C. Elmore and M.A. Heald. *Physics of waves* (Dover Publications, New York, 1969)
- [3] E. Orsingher and M. D'Ovidio. Pseudo-processes related to the equation of vibrations of rods (in preparation)

## Analyzing spatial circular data using wrapped Gaussian processes

Giovanna Jona Lasinio, Alan Gelfand, Mattia Jona Lasinio

**Keywords:** Bayesian kriging, Circular data, wrapped distribution

### Abstract

Periodical aka circular data arise in several context, such as oceanography (wave directions) and meteorology (wind directions), and, more generally, with measurements on a circular scale. This study introduces a fully model-based approach to handle circular data in the case of measurements taken at spatial locations, anticipating structured dependence between these measurements. We formulate a wrapped Gaussian spatial process model for this setting, induced from a customary inline Gaussian process. We build a hierarchical model to handle this situation and show that the fitting of such model is possible using standard Markov chain Monte Carlo methods. Our approach enables spatial interpolation and can accommodate measurement error. This study is motivated by the need of downscaling a set of wave direction data from the Adriatic coast of Italy. Data are generated through a complex computer model (WAM). These data are produced on a grid with 25X25 km cells and are related to wave height forecast in deep waters (more the 100m depth). In order to include directional information in shallow waters forecast models it is necessary to align these values with other data given at a higher spatial resolution (10X10 Km grid).

## **Data Augmentation Approach in Bayesian Modelling of Presence-only Data**

F. Divino, N. Golini, G. Jona Lasinio, A. Penttinen

**Keywords:** Bayesian model, Data augmentation, Presence only-data

### Abstract

An important issue in ecological studies is to estimate the potential spatial extent of an ecological niches. The prediction of geographic distribution of species in suitable areas is essential for planning conservation and management strategies, and it may concern animals and plants species studies. Given the presence-absence data for a species, a logistic model represents a natural approach to estimate the species distribution, given the environmental covariates. In ecological studies, however, collecting such presence-absence data could be expensive and/or difficult. Often the only available information is the true presence of the species and the associated environmental covariates in some locations of the study area. Such data are known in literature as presence-only data, see Pearce and Boyce (2006). Following the pseudo-absence approach (Ward et al., 2009), it is possible to adapt standard case-control models to presence-only data. We propose a Bayesian model to estimate logistic linear regressions adapted to presence-only data of rare species. We introduce a random approximation of the correction factor in the adjusted logistic model that allows us to overcome the need to know the prevalence of the species. Under the assumption that the pseudo-absence locations are randomly sampled from the entire study area, we can estimate regression parameters jointly with prevalence through a data augmentation MCMC algorithm (Liu, 2004). Preliminary simulations investigating the behaviour of our model with respect to the spatial structure of the covariates, to the rareness of the species and to the dimension of the pseudo-absences sample have shown encouraging results. References: - Liu S. J. (2004), "Monte Carlo Strategies in Scientific Computing", Springer, New York. - Pearce, J. L. and Boyce, M. S. (2006), "Modelling distribution and abundance with presence-only data", *Journal of Applied Ecology*, 43, 405-412. - Ward, G., Hastie, T., Barry, S., Elith, J. and Leathwick, A. (2009), "Presence-only data and the EM algorithm", *Biometrics*, 65, 554-563.

## **Exchange Rate and Wage Adjustment**

Francesco Nucci - Sapienza Università di Roma  
and  
Alberto Franco Pozzolo - Università del Molise}

In this paper we rely on a representative panel of manufacturing firms and estimate the implications for firm-level wages of changes in the permanent component of the exchange rate. Similarly to the response of employment and hours documented in Nucci and Pozzolo (JIE, 2010), the direction and size of wage adjustment is shaped by the external orientation of each firm on both the sale and cost side of the balance sheet.

In the first section we present a simple model of the labor market that allows us to illustrate the mechanisms through which exchange rate swings induce equilibrium wage adjustment. Following Nucci and Pozzolo (2001; 2010) and Campa and Goldberg (2001), we consider the optimal conditions for profit maximization of a firm operating in an imperfectly competitive market. The theoretical model allows us to derive an expression for the elasticity of real wages with respect to exchange rate. This equation represents a useful theoretical background for our empirical analysis, providing a variety of testable implications that lend themselves to the data scrutiny. We discuss in the paper the number of channels, explicitly identified in our theoretical model, through which exchange rate movements affect wages.

Our model predicts that exchange rate swings affect wage adjustments. In particular, if we focus on the revenue side of the firm's balance sheet a currency depreciation has a positive effect on the firm's marginal revenue product of labor and therefore on real wages. In particular, a depreciation is predicted to exert a positive effect on wages along the channels of both foreign and domestic sales. By contrast, a currency depreciation negatively affects, through the cost side of the balance sheet, the marginal revenue product of labor and thereby real wages. Very importantly, the extent of these adjustments depends on the external orientation of the firm towards international product markets. On the export side, the positive effect on wages of an exchange rate depreciation is larger the higher is the share of foreign sales on total sales, the more a firm is exposed to foreign markets through the export of its products. On the revenue side as a whole, i.e. including both domestic and foreign sales, the positive effect on wages on a currency depreciation is also amplified as the export share of sales becomes larger, but this prediction holds true only if the sum of the exchange rate pass-through elasticities (their absolute value) is less than one (Nucci and Pozzolo, JIE 2010). On the expenditure side, conversely, the size of the negative effect on wages of a currency depreciation depends on the extent of firm's reliance upon imported inputs. This type of exposure is captured in the theoretical equation by the share of expenditure on imported inputs on total costs: the larger is this share is, the more pronounced the negative implications for wages of a depreciation through this channel of exposure.

The theoretical framework highlights other important features that shape the response of wage to exchange rate swings. First, the firm's degree of market power, as measured by the firm's mark-up ratio, does enters the equation and the prediction is that, everything else being identical for two firms (e.g., type of international exposure and pass-through elasticities), the one with a lower degree of monopoly power tends to exhibit a more sizeable (absolute value of the) elasticity of wage to exchange rate. Indeed, firms with a lower market power tend to be less capable to absorb currency shocks so that the impact on wages is relatively more pronounced. To see this, recall that an exchange rate depreciation drives down the export price in the foreign currency by an amount that depends on the pass-through elasticity. This price decline therefore yields an increase of foreign demand and thereby of profitability and wages, which is larger the higher is the price elasticity of

foreign demand. Given the negative relationship between the firm's mark-up ratio and the price elasticity of demand, the sensitivity of wage to currency swings is indeed magnified when the firm's market power is relatively low.

Moreover, another prediction of the model is that the exchange rate pass-through elasticities contribute to shape the adjustment of wage in response to currency shocks. The exchange rate elasticity of prices set by the firm in the destination market's currency ranges from zero (no pass-through) to one (complete pass-through). We also note that, for a given level of external orientation on the export side, the smaller is the elasticity of exchange rate pass-through to foreign prices, the larger is the (absolute value of the) wage response to a shift in exchange rate. Importantly, a number of contributions show that this exchange rate pass-through elasticity does depend on market structure and in particular on the extent to which firms' products are differentiated and the substitution among different variants is large (Yang, 1997). These studies include Dornbusch (1987) and Knetter (1993) and show that the pass-through tends to be low if the degree of competition in the foreign markets is high. Therefore, in the limiting case of a perfectly competitive foreign destination market, the firm is a price taker and the pass-through elasticity would be zero. Thus, the prediction through this channel contributes to reinforce the previous conclusion that the lower is the firm's pricing power the higher is the exchange rate sensitivity of wages.

If we focus on firm's competition in the domestic market, we have established that a currency depreciation, by making foreign products more expensive, rises the competitiveness of domestic firms in the home market, thus increasing their sales and thereby their profitability and wages. Our prediction is that the exchange rate pass-through elasticity of firm's prices in the domestic market plays an important role in shaping this effect. The values of this domestic pass-through elasticity range from minus one (complete pass-through) to zero (no pass-through), and again it depends on market structure. Specifically, the elasticity is (in absolute value) a decreasing function of the firm's monopoly power in the home market. In the limiting case of a perfectly competitive domestic destination market, where the domestic firm is a price taker, a currency appreciation must be rebated by a one-to-one firm's own price cut (a pass-through elasticity equal to minus one). Thus, a currency appreciation induces a reduction in the value of domestic sales, profitability and wages and the lower the firm's market power the larger this decline turns to be. The intuition is that the higher the competitive pressure exerted by foreign producers, the more responsive domestic sales, profitability and wages are in the aftermath of a currency shift. Indeed the domestic pass-through elasticity reflects the degree of this competitive pressure from foreign producers and is often assumed to be proportional to the degree of import penetration in the domestic market (see e.g. Dornbusch, 1987 and Campa and Goldberg, 2001).

Market structure also affects the effects through the expenditure side of the balance sheet. Indeed, the increase of wages after an exchange rate appreciation taking place through a reduction in the expenditure for imported inputs depends on the extent of competition in the market for intermediate inputs. Indeed, we establish that the effect of exchange rate on wage adjustment is larger the smaller the pass-through elasticity of foreign input prices to the exchange rate. The latter ranges from zero (no pass-through) to one (complete pass-through).

In the empirical investigation we rely on microeconomic data drawn from two different statistical sources. The first one is the Bank of Italy's Survey of Investment in Italian Manufacturing (SIM) carried out at the beginning of every year since 1984 on a representative sample of over 1,000 firms stratified by industry, firm size and location. The data are of extremely high quality also for the professional expertise of the interviewers, who are officials of the Bank of Italy establishing long-term relationships with the firms' managers. In order to ensure the quality standard of data, only medium-large firms, defined as those with more than 50 employees, are included in the Survey. We

use SIM to gather firm-level information on total revenues and revenues from exporting, employment and hours worked. The second data source of firm level data is the Company Accounts Data Service reports, a database maintained by a consortium of the Bank of Italy and a pool of banks, collecting information from balance sheets and income statements of a sample of about 40,000 Italian firms. The detailed information from the annual accounts are reclassified to ensure comparability across firms. This database provides us with firm level information on labor compensation, intermediate input expenditure, value added and gross output. We merged the data from the two sources to construct an unbalanced panel of slightly fewer than 2,400 firms. As in Nucci and Pozzolo (2010) the data used for estimation covers the period 1984-1998, the one antecedent to the introduction of the Euro.

For measuring exchange rate, we separately use both the export and import real effective exchange rates of the Italian lira constructed by considering 24 different bilateral exchange rates. Both real exchange rates are computed using producer price indexes (see Banca d'Italia, 1998). In the empirical analysis, we use the permanent component of exchange rate variations which has been derived by applying the Beveridge and Nelson (1981) procedure that decomposes a non-stationary series into its permanent and transitory components (see Nucci and Pozzolo, 2010 for details).

In assessing on empirical grounds the wage response to exchange rate fluctuations we rely on a baseline specification which can be seen as the empirical counterpart of the theoretical equation and allows us to test a number of theoretical implications of the model. The advantage of this approach is that the estimated sensitivity of wages to currency movements varies across firms and over time depending on the evolving external orientation of each specific firm on both the revenue and cost side. In other words, the empirical framework allows us to derive a time-varying firm specific estimated response of firms' labor compensation to exchange rate oscillations.

The methodology used for estimation is the generalized method of moments (GMM) estimator for dynamic panel data model, which was shown to be efficient within the class of instrumental variable estimators (Arellano and Bond, 1991). Indeed, since in the estimating equation the lagged values of the mark-up ratio and of change in employment (hours) and sales are likely to be correlated with the firm-specific fixed effects, this endogeneity of regressors would cause inconsistency of the parameters estimated with standard panel methods while the GMM estimator would ensure their consistency. Specifically, following Arellano and Bover (1995) and Blundell and Bond (2000), we rely on the system GMM panel estimator which augments the Arellano and Bond (1991) estimator by building a system of two equations: the original equation and a transformed one. As in Arellano and Bond (1991), in the transformed equation a variety of instruments in levels can be used. However, under the novel approach a further assumption is made: that first differencing the instrumenting variables in the original equation makes them uncorrelated with fixed effects. This allows us to exploit an even larger number of orthogonality conditions than before, by resorting to a larger instrument set.

Our empirical results are the following. Through the revenue side, wages are shown to rise after a currency depreciation and the effect is estimated to be larger the higher is the firm's exposure to sales from exports. On the contrary, a depreciation induces a cut in firm's wages through the expenditure side of the balance sheet and the effect is larger the higher is the incidence of imported inputs on total production costs. We show that, for a given degree of firm's external orientation, the sensitivity of wages to exchange rate swings is larger for firms with a lower market power. We also show that other transmission channels introduce a degree of specificity in the effect of exchange rate on wages. These include the degree of sectoral import penetration in the domestic market, the extent of input's substitutability in production and the composition by type of firm's labor force.

# Moto Browniano Iperbolico su Semispazi di Poincaré

Valentina Cammarota <sup>†</sup>    Enzo Orsingher <sup>‡‡</sup>

Per semplicità consideriamo il semipiano di Poincaré (modello di spazio non Euclideo Iperbolico)

$$\mathbb{H}^2 = \{(x, y) : x \in \mathbb{R}, y > 0\}$$

dotato della metrica

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

Su di esso consideriamo un moto Browniano iperbolico, cioè una diffusione con generatore in coordinate cartesiane

$$\frac{y^2}{2} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)$$

e in coordinate iperboliche

$$\frac{1}{2} \left( \frac{\partial^2}{\partial \eta^2} + \frac{1}{\tanh \eta} \frac{\partial}{\partial \eta} + \frac{1}{\sinh^2 \eta} \frac{\partial^2}{\partial \alpha^2} \right).$$

Si dimostra che per un moto Browniano che parte dal punto  $z = (\eta, \alpha)$  la probabilità di colpire un cerchio iperbolico di centro  $(0, \cosh \bar{\eta})$  e di raggio  $\sinh \bar{\eta}$  è data da

$$\mathbb{P}_z\{B_{\mathbb{H}^2}(T_{\bar{\eta}}) \in d\bar{\alpha}\} = \frac{1}{2\pi} \frac{\cosh \bar{\eta} - \cosh \eta}{\cosh \eta \cosh \bar{\eta} - 1 - \sinh \eta \sinh \bar{\eta} \cos(\alpha - \bar{\alpha})} d\bar{\alpha}, \quad \bar{\alpha} \in [0, 2\pi). \quad (0.1)$$

Per  $\eta$  and  $\bar{\eta}$  piccolo dalla (0.1) si ricava il classico nucleo di Poisson Euclideo che fornisce

$$\mathbb{P}_z\{B_{\mathbb{R}^2}(T_{\bar{\eta}}) \in d\bar{\alpha}\} = \frac{1}{2\pi} \frac{\bar{\eta}^2 - \eta^2}{\bar{\eta}^2 + \eta^2 - 2\eta\bar{\eta} \cos(\alpha - \bar{\alpha})} d\bar{\alpha}, \quad \bar{\alpha} \in [0, 2\pi).$$

Per  $\bar{\eta} \rightarrow \infty$  dalla (0.1) si estrae la legge di Cauchy la cui densità è, in coordinate cartesiane,

$$\mathbb{P}_z\{B_{\mathbb{H}^2}(T_{\infty}) \in d\bar{x}\} = \frac{1}{\pi} \frac{\frac{y}{x^2+y^2}}{\left[\bar{x} - \frac{x}{x^2+y^2}\right]^2 + \left[\frac{y}{x^2+y^2}\right]^2} d\bar{x},$$

mentre in coordinate iperboliche diventa

$$\mathbb{P}_z\{B_{\mathbb{H}^2}(T_{\infty}) \in d\bar{\alpha}\} = \frac{1}{2\pi} \frac{1}{\cosh \eta - \sinh \eta \cos \alpha \cos \bar{\alpha} - \sinh \eta \sin \alpha \sin \bar{\alpha}} d\bar{\alpha}.$$

<sup>†</sup>Dipartimento di Statistica, Sapienza Università di Roma, P.le Aldo Moro 5, 00185 Rome, Italy. Tel.: +390649910499, fax: +39064959241. E-mail address: valentina.cammarota@uniroma1.it.

<sup>‡‡</sup>*Corresponding author.* Dipartimento di Statistica, Sapienza Università di Roma, P.le Aldo Moro 5, 00185 Rome, Italy. Tel.: +390649910585, fax: +39064959241. E-mail address: enzo.orsingher@uniroma1.it.

Il moto Browniano sulla sfera  $S^2$  con punto di partenza in  $z = (\theta, \phi)$  colpisce la calotta polare di latitudine  $\bar{\theta}$  con una distribuzione

$$\mathbb{P}_z\{B_{S^2}(T_{\bar{\theta}}) \in d\bar{\phi}\} = \frac{1}{2\pi} \frac{\cos \theta - \cos \bar{\theta}}{1 - \cos \theta \cos \bar{\theta} - \sin \theta \sin \bar{\theta} \cos(\phi - \bar{\phi})} d\bar{\phi}. \quad (0.2)$$

L'analogia tra il risultato (0.1) e (0.2) è dovuta al fatto che il semipiano di Poincaré è una sfera di raggio immaginario e quindi il ruolo della geometria sferica viene rimpiazzato dalle funzioni iperboliche della trigonometria iperbolica. Rilevanti differenze tra moto Browniano iperbolico e euclideo si possono percepire studiando le probabilità di uscita da una corona limitata da due sfere concentriche di raggi  $\sinh \eta_1$  e  $\sinh \eta_2$ . Per

$$T_{\eta_j} = \inf(t > 0 : \eta(O, B_{\mathbb{H}^2}(t)) = \eta_j)$$

si ricava che

$$\mathbb{P}_z\{T_{\eta_1} < T_{\eta_2}\} = \frac{\log \tanh \frac{\eta_2}{2} - \log \tanh \frac{\eta}{2}}{\log \tanh \frac{\eta_2}{2} - \log \tanh \frac{\eta_1}{2}}, \quad \eta_1 < \eta < \eta_2$$

e per  $\eta_2 \rightarrow \infty$  fornisce

$$\mathbb{P}_z\{T_{\eta_1} < \infty\} = \frac{\log \tanh \frac{\eta}{2}}{\log \tanh \frac{\eta_1}{2}}, \quad \eta_1 < \eta, \quad (0.3)$$

che mostra che con probabilità positiva si può evitare un cerchio iperbolico per un tempo infinito (nel caso euclideo) si ha invece

$$\mathbb{P}_z\{T_{\eta_1} < \infty\} = 1$$

come si deduce dalla espressione che si ricava da (0.3) per  $\eta$  piccolo, cioè

$$\mathbb{P}_z\{T_{\eta_1} < T_{\eta_2}\} \sim \frac{\log \eta_2 - \log \eta}{\log \eta_2 - \log \eta_1}, \quad \eta_1 < \eta < \eta_2.$$

Gran parte del lavoro è dedicata allo studio della probabilità di colpire ipersfere iperboliche da parte del moto Browniano iperbolico  $n$ -dimensionale, cioè la diffusione con generatore

$$\Delta_n = \frac{y^2}{2} \left( \sum_{i=1}^{n-1} \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y^2} \right) - \frac{(n-2)}{2} y \frac{\partial}{\partial y}.$$

I casi  $n = 2$  e  $n > 2$  sono qualitativamente differenti e nel secondo le leggi di probabilità ottenute sono espresse mediante funzioni ipergeometriche e polinomi di Gegenbauer. Per  $\eta$  piccolo si ricava il nucleo di Poisson Euclideo che possiede la forma

$$\frac{\Omega_{n-1}}{\Omega_n} \frac{1 - \frac{\eta^2}{\bar{\eta}^2}}{\left(1 - \frac{2\eta}{\bar{\eta}} \cos(\alpha_1 - \bar{\alpha}_1) + \frac{\eta^2}{\bar{\eta}^2}\right)^{\frac{n}{2}}}$$

dove  $\Omega_n = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}$  è l'area della superficie dell' ipersfera  $S^n$ .

## References

- [1] Baldi, P., Casadio Tarabusi, E., Figà-Talamanca, A., Yor, M.: Non-symmetric hitting distributions on the hyperbolic half-plane and subordinated perpetuities. *Rev. Mat. Iberoam.*, **17**, 587–605 (2001)

- [2] Byczkowski, T., Graczyk, P., Stós, A.: Poisson kernels of half-spaces in real hyperbolic spaces. *Rev. Mat. Iberoam.*, **23**, no. 1, 85–126 (2007)
- [3] Byczkowski, T., Malecki, J.: Poisson kernel and Green function of the ball in the real hyperbolic spaces. *Potential Anal.*, **27**, 1–16 (2007)
- [4] Cammarota, V., Orsingher, E.: Travelling randomly on the Poincaré half-plane with a Pythagorean compass. *J. Stat. Phys.*, **130**, 455–482 (2008)
- [5] Cammarota, V., Orsingher, E.: Cascades of Particles Moving at Finite Velocity in Hyperbolic Spaces. *J. Stat. Phys.*, **133**, 1137–1159 (2008)
- [6] Chavel, I.: Eigenvalues in Riemannian Geometry. Academic Press, New York, London (1984)
- [7] Davies, E., B.: Heat kernels and spectral theory. Cambridge Tracts in Mathematics, **92**. Cambridge University Press, Cambridge (1990)
- [8] Grigor'yan, A.: Analytic and geometric background of recurrence and non-explosions of the Brownian motion on Riemannian manifolds. *Bull. American Math. Soc.* **36**, 135–249 (1999)
- [9] Gruet, J-C.: A note on hyperbolic von Mises distributions. *Bernoulli*, **6**, no. 6, 1007–1020 (2000)
- [10] Helgason, S.: Groups and Geometric Analysis. Academic Press, New York, London (1984)
- [11] Lao, L., Orsingher, E.: Hyperbolic and fractional hyperbolic Brownian motion. *Stochastics*, **79**, no. 6, 505–522 (2007)
- [12] Matsumoto, H., Yor, M.: Exponential functionals of Brownian motion. I. Probability laws at fixed time. *Probab. Surv.* **2**, 312–347 (2005)

## Statistica a scuola: strategie per un insegnamento efficace

M. Gabriella Ottaviani

Nel marzo 2010, il regolamento di riordino degli Istituti professionali, quello degli istituti tecnici e quello dei licei hanno inserito statistica e probabilità nei curricula delle scuole superiori (o per le meno al primo biennio) sicché gli insegnanti delle prime classi delle superiori si sono trovati alle prese dal settembre 2010 con l'insegnamento di Statistica e probabilità, o meglio di "Dati e previsioni". Diversa la situazione per la scuola di base dove l'introduzione della statistica nei programmi di insegnamento nella scuola di base italiana risale al 1979 per la scuola media (oggi scuola secondaria di primo grado) e al 1985 per la scuola elementare (oggi scuola primaria) ed è stata rivisitata nel 2004, con vari rimaneggiamenti nel 2007, senza però modificarne di fatto i contenuti.

A questo punto si potrebbe pensare di aver raggiunto una situazione ottimale: la statistica e la probabilità sono dappertutto, si tratta di mietere ciò che si è seminato!!! Nulla di più falso. Anche se il Presidente della CIIM, l'organismo dell'UMI che si occupa di didattica della Matematica, sottolineava "l'ampliamento di orizzonti" (Bolondi, 2010) che, "(ridisegnando) in maniera netta gli obiettivi di apprendimento" (Bolondi, 2010) della nostra scuola secondaria superiore, ha finalmente generalizzato a tutte le tipologie di scuole superiori l'insegnamento della statistica, ciò in realtà crea nella scuola, in particolare tra i docenti, una notevole quantità di problemi pratici.

Il percorso di studi di una larga parte degli insegnanti di matematica infatti non prevede corsi di statistica, né tanto meno di didattica della statistica;

Il non aver generalmente ricevuto un insegnamento di statistica e probabilità è una carenza grave per gli insegnanti della nostra scuola secondaria che nella loro larga maggioranza sono diventati insegnanti conoscendo al più la disciplina che insegnano, mentre dal punto di vista della didattica si rifanno al migliore fra gli insegnanti che hanno avuto la sorte di incontrare. Nel sistema di istruzione italiano non è, di fatto, al momento ufficialmente implementato un percorso di formazione professionale iniziale per gli insegnanti della secondaria. Ciò rende particolarmente difficile far entrare in classe argomenti innovativi come probabilità e statistica. Non essendoci tradizione di insegnamento, l'insegnante di matematica ha nella generalità dei casi due lacune da colmare: il contenuto disciplinare di statistica e probabilità ed il modo di porgerli (Ottaviani, 2009).

Inoltre, per un insegnante, è importante l'opinione che egli ha degli argomenti che insegna. Poiché capita sovente che chi non conosce un argomento ne ha una opinione non positiva, ciò tende a creare, soprattutto per statistica, un circolo vizioso che può essere rotto, facendo conoscere agli insegnanti la disciplina, in modo che conoscendola, anche nei suoi rapporti con la probabilità e con gli altri nuclei della matematica, ne colgano l'utilità, trovino gusto al suo insegnamento e sappiano implementarlo in classe.

Inoltre, anche se statistica e probabilità, nell'insegnamento di matematica a scuola appartengono allo stesso nucleo: "Dati e previsioni", in effetti non hanno la stessa natura; ciò deve emergere in modo da potenziare l'insegnamento di entrambe. Se la probabilità ha il compito di affrontare lo studio dell'incertezza, essa appartiene alla matematica e, come il ragionamento matematico, è deduttivo (Anichini, 2010). La statistica invece ha lo scopo di studiare in modo scientifico i fenomeni collettivi, ossia quei fenomeni che si possono conoscere solo eseguendo una massa di osservazioni (Gini, 1962). La probabilità è la cerniera naturale fra la matematica e la statistica,

ed entrambe guardano da punti di vista diversi la realtà sensibile: l'una per studiare la variabilità delle popolazioni, l'altra per studiare l'incertezza di singoli eventi. Può allora la probabilità da sola assumersi la responsabilità e sostenere il carico di fare da traino all'inserimento della statistica all'interno dell'insegnamento della matematica a scuola?

In effetti, nella scuola di secondo grado un certo malessere si avverte nei confronti della statistica e non solo oggi, e non solo in Italia, dato che si osserva ovunque nel mondo una mancanza di preparazione di base in statistica nella formazione degli insegnanti. Ma a ben guardare (Gattuso, Ottaviani, 2011) le differenze fra il modo di ragionare della matematica e della statistica, se pur diversi, sono fra loro complementari.

Tra i punti principali di diversità della statistica dalla matematica si segnalano:

- *la diversità nel tipo di ragionamento: quello statistico è induttivo, quello matematico è deduttivo;*
- *la diversa importanza attribuita al contesto: la statistica insiste sulla interpretazione del contesto, mentre la matematica promuove l'astrazione;*
- *la diversità con la quale entrambi guardano ai numeri;*
- *la diversità con cui entrambe considerano la variabile e la sua misura .*

Questi elementi di diversità sono però anche quelli su cui far leva per favorire un insegnamento della matematica che non abbia come scopo solo l'acquisizione di risultati e di tecniche. Come diceva Giovanni Prodi (1984) l'insegnamento /apprendimento della matematica a scuola deve educare con gradualità gli studenti a pensare. Senza dubbio gli elementi di Statistica e probabilità inseriti negli ordinamenti possono aiutare proprio a questo. Se la scuola Italiana fino a qualche anno fa ha mostrato di dare al più contenuti, ma di non aiutare i propri studenti a sviluppare competenze (cfr esiti di PISA 2000), le nuove indicazioni per le scuole di tutti i livelli danno valore a sapere, saper fare e ai processi posti in essere dagli studenti per l'acquisizione della propria competenza matematica. D'altro canto, esistono progetti (Matematica del cittadino, Piano m@t.abel) che hanno proposto un insieme di attività didattiche costruite tenendo conto delle capacità di intuizione che nascono quando l'insegnamento della matematica è calata in un contesto, in modo che la necessità della matematica nasca dall'esigenza di risolvere problemi e la formalizzazione sia sentita come una necessità anziché essere imposta a priori. La statistica è in grado di dare valore e sostanza a questo modo laboratoriale di porgere la matematica. Ci si augura che questo momento favorevole al potenziamento dell'insegnamento/apprendimento della matematica, aiuti finalmente la statistica ad entrare a scuola come una disciplina essenziale, anche per la crescita scientifica della nostra cultura.

Con ogni probabilità ciò richiederà ulteriori ricerche non solo su "come" e "cosa" proporre in classe, ma anche sull'individuazione dei momenti migliori per inserire con efficacia statistica e probabilità nella programmazione scolastica, così da svilupparne ed equilibrarne l'insegnamento insieme con le altre parti della matematica. Il supporto della ricerca in didattica della statistica sarà perciò ancora indispensabile per aiutare l' insegnamento/apprendimento della disciplina a scuola.

## Bibliografia

ANICHINI, G., *Matematica e statistica: differenze, contatti e...connivenze, Dossier Treccani 2010:*

*La statistica a scuola*

<http://www.treccani.it/Portale/sito/scuola/dossier/2010/statistica/anichini.html>

BOLONDI, G., *E' questa la riforma della matematica che manda in pensione Gentile*, 25 marzo 2010, [http://dirispl.interfree.it/riforma/bolondi-riforma\\_matematica.htm](http://dirispl.interfree.it/riforma/bolondi-riforma_matematica.htm)

- GINI, C., *La logica della statistica*, Boringhieri, Torino, 1962.
- OTTAVIANI, M. G., Ho insegnato statistica alla SSIS del Lazio, *Induzioni*, 39, 2009, pp. 85-97
- PRODI, G., *Matematica come scoperta: per il biennio delle scuole superiori*, 1, Seconda edizione ampliata, Firenze, D'Anna, 1984.
- GATTUSO, L., OTTAVIANI, M. G., Complementing mathematical thinking and statistical thinking in school mathematics. In Batanero, C, Burrill, G. Reading, C. (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education*, A Joint ICMI/IASE Study, New ICMI Study Series, Vol. 15, Springer, ISBN: 978-94-007-1130-3. Previsto per agosto 2011

Paolo Palazzi, *L'economia come scienza sociale e politica*, Aracne ed , 2010

Ho raccolto alcuni recenti scritti non pubblicati che rispecchiano e descrivono vari miei interessi culturali. Ho raggruppato gli scritti in cinque capitoli, ognuno dei quali rappresenta un aspetto della mia vita culturale e professionale:

1) *La teoria economica*. Lo studio della teoria economica mi ha lasciato sempre più confuso e perplesso mano a mano che si ampliavano le mie letture e mano a mano che tentavo di seguire l'evolversi della disciplina economica. Si potrebbe descrivere il mio atteggiamento culturale nei confronti della teoria economica con un grafico a U rovesciato, dove nelle ordinate c'è il livello di interesse e la curiosità di conoscere l'evolvere della teoria economica e nelle ascisse gli anni di studio.

Mano a mano che studiavo e riuscivo a capire l'approccio teorico neoclassico, il mio calo di interesse per la teoria economica diminuiva in proporzione. Più studiavo e più mi allontanavo dall'interesse per lo studio della teoria economica. Nel migliore dei casi mi sembravano esercizi di logica matematica, nella maggioranza dei casi inutili e dannosi tentativi di costruire mondi perfetti. L'economia intesa come strumento per la comprensione della realtà era letteralmente sparita.

In questo capitolo del volume sono raccolti tre interventi, il primo relativo alla mia critica generale, vicina allo sfogo, della deriva che la teoria economica negli ultimi decenni ha preso; degli altri due, il primo descrive uno schema generale dal quale, a mio avviso, è possibile partire per portare di nuovo la teoria economica a interessarsi della interpretazione della realtà, ritornando a essere considerata una scienza sociale e politica e quindi con un ampio uso della multidisciplinarietà; il secondo scritto è relativo al rapporto chiave tra ecologia ed economia, rapporto a mio avviso in grado di essere affrontato proficuamente soltanto da una teoria economica che abbandoni quasi completamente le impostazioni teoriche *mainstream*.

2) *La politica*. I primi scritti proposti in questo capitolo rispecchiano l'evolversi di questi miei pensieri e convinzioni, gli ultimi sono invece due proposte di intervento legislativo frutto di discussioni e lavoro con un gruppo di amici in un periodo di collaborazione con il movimento ecologista.

3) *L'economia italiana*. Spesso gli economisti accademici non si interessano alla realtà, o meglio si costruiscono una loro realtà teorica e, se scoprono che la realtà reale funziona in modo diverso, considerano sbagliata quest'ultima e non la loro teoria. Gli articoli proposti in questo capitolo si occupano essenzialmente di tre aspetti: occupazione, sviluppo sostenibile e tassazione. L'ultimo scritto è invece relativo alla recente crisi finanziaria.

4) *Lo sviluppo e il sottosviluppo*. In questo capitolo sono riportate alcune idee generali sulle teorie del sottosviluppo. Partendo dal presupposto di considerare come fenomeno negativo lo sviluppo diseguale e illustrare le teorie e le politiche che hanno tentato e tentano di spiegare il sottosviluppo e propongono soluzioni più o meno semplici per superarlo, tenere un corso sul problema del sottosviluppo è molto frustrante. Infatti siamo in una realtà in cui il divario economico e sociale fra paesi e fra persone si allarga continuamente, le politiche economiche nazionali e internazionali sperimentate sono state di tutti i tipi, anche contrastanti, ma tutte accomunate da drammatici fallimenti.

Ancor più frustrante è stato per me insegnare queste teorie in una facoltà come Scienze statistiche, nella quale gli studenti erano da sempre abituati a porsi problemi e dubbi, ma, attraverso il corretto utilizzo di tecniche matematiche e statistiche,

ritenevano sempre possibile raggiungere una soluzione soddisfacente. Invece al termine delle mie lezioni le domande senza risposta erano più numerose rispetto a quelle poste all'inizio delle lezioni stesse.

Alla domanda finale degli studenti: «Quali possono essere le soluzioni?», la mia risposta era: «Non lo so, speriamo che qualcuno di voi riesca a trovarle». Facce attonite della maggioranza degli studenti e un evidente senso di delusione e forse di frustrazione per aver seguito delle lezioni che, a parte l'ovvio e comunque rassicurante obiettivo della maggiore possibilità del superamento dell'esame, non li hanno sufficientemente rassicurati con convincenti soluzioni del problema affrontato. Questo livello di incertezza e di dubbi, non esprimibili e in genere non accettati nell'accademia, è invece presente negli scritti proposti in questo capitolo.

5) *La scuola*. Questo ultimo capitolo raccoglie quattro brevi interventi sulla scuola, due dedicati al rapporto tra scuola pubblica e privata, due relativi alla mia relazione con la scuola come docente e come genitore di studenti.

## Indice

Premessa .....	7
Capitolo 1 <i>La teoria economica</i>	
1.1 Economia: scienza divina? .....	13
1.2 Uno schema di produzione e riproduzione .....	16
1.3 Problemi ecologici e sviluppo .....	42
Capitolo 2 <i>La politica</i>	
2.1 Politica e morale .....	57
2.2 Berlusconismo e idee di sinistra .....	60
2.3 Quale politica per un partito ecologista? .....	63
2.4 L'alfabetizzazione multimediale e le 150 ore .....	72
2.5 Le unioni solidali .....	79
2.6 Legalità e classismo .....	93
2.7 Emigrazione, migrazione o invasione? .....	96
2.8 Riflessioni sparse ed "esagerate" sulle guerre "moderne" .....	99
Capitolo 3 <i>L'economia italiana</i>	
3.1 Politica per uno sviluppo sostenibile .....	105
3.2 La globalizzazione come sfida per una nuova qualità della vita .....	122
3.3 Occupazione e qualità della vita: come? Una proposta di istituzione di un Salario di Attività Sociale (SAS) .....	130
3.4 Se 35 ore vi sembrano poche .....	138
3.5 La formazione di capacità imprenditoriale .....	141
3.6 Il problema fiscale .....	155
3.7 Aumentare le tasse o ridurre la spesa pubblica? .....	158
3.8 Debito pubblico: perché no? .....	169

3.9	In “difesa” di banche e agenzie finanziarie .....	174
-----	---------------------------------------------------	-----

Capitolo 4 *Lo sviluppo e il sottosviluppo*

4.1	Appunti sulle teorie economiche dello sviluppo diseguale .....	179
4.2	Le politiche degli aiuti ai paesi sottosviluppati .....	229
4.3	Aiuti internazionali e democrazia .....	243
4.4	Fame e democrazia .....	246
4.5	Riflessioni di un economista sul rapporto tra crescita demografica e sviluppo economico .....	250
4.6	La socialità e la nuova povertà urbana .....	269
4.7	Terrorismo, povertà e Tobin tax .....	276

Capitolo 5 *La scuola*

5.1	Il finanziamento pubblico della scuola privata .....	279
5.2	Lettera a due studenti durante l’occupazione della loro scuola .....	283
5.3	Una giornata di un professore universitario .....	286

## **Sulla ottimizzazione di trattati di riassicurazione mediante tecniche di “Parallel Tempering”**

**Luca Passalacqua**

Nel quadro della normativa “Solvency II”, verranno presentate le problematiche relative alla ottimizzazione di trattati di riassicurazione non proporzionali per l’assicurazione danni.

In particolare si farà riferimento al ramo 14 – assicurazione del credito – e ai risultati ottenuti in [Pa-09] dove tale problema è stato risolto tramite la tecnica del “*simulated annealing*” [KG-83] nella sua versione adattiva proposta da Ingber [In-96].

Successivamente si discuterà di come sia possibile migliorare, dal punto di vista della precisione e del tempo di calcolo, la soluzione del problema di ottimo impiegando in ambiente di calcolo parallelo la tecnica nota come “*parallel tempering*” [Ge-91, HN-96].

Infine si farà riferimento a possibili implementazioni a basso costo tramite schede grafiche NVIDIA e il linguaggio CUDA [KH-10].

### **Riferimenti bibliografici**

- [Pa-09] L. Passalacqua, “Optimal trade credit insurance programs with Solvency requirements”, *Giornale dell’Istituto Italiano degli Attuari*, LXXII, n. 1-2, 2009, pp. 147-176.
- [In-96] L. Ingber, “Adaptive Simulated Annealing (ASA): Lessons learned”, *Journal of Control and Cybernetics* 25, 1996, pp. 33-54.
- [KG-83] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, “Optimization by Simulated Annealing”, *Science*, 220, 4598, pp.671-680.
- [Ge-91] C.J. Geyer, in “Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface”, Ed. E.M.Keramidas (Interface Found. Amer., Fairfax Station, VA, 1991) pp.156-163.
- [HN-96] K. Hukushima and K. Nemoto, “Exchange Monte Carlo Method and Application to Spin Glass Simulations”, *Journal of the Physical Society of Japan*, 65 n. 6, 1996, pp. 1604-1608.
- [KH-10] D. Kirk, W-M. Hwu, “Programming massively parallel processors”, Morgan Kaufmann, Burlington, 2010.

# Searching with Memory Faults: An Experimental Study

Umberto Ferraro-Petrillo<sup>1</sup>, Fabrizio Grandoni<sup>2</sup>, and Giuseppe F. Italiano<sup>2</sup>

<sup>1</sup>Dipartimento di Scienze Statistiche, Sapienza Università di Roma, P.le Aldo Moro 5, 00185 Rome, Italy. Email: [umberto.ferraro@uniroma1.it](mailto:umberto.ferraro@uniroma1.it). <sup>2</sup>Dipartimento di Informatica, Sistemi e Produzione, Università di Roma Tor Vergata, Via del Politecnico 1, 00133 Roma, Italy. Email: {grandoni,italiano}@disp.uniroma2.it.

**Abstract.** We address the problem of implementing data structures resilient to memory faults which may arbitrarily corrupt memory locations. In this framework, we focus on the implementation of dictionaries, and perform a thorough experimental study using a testbed that we designed for this purpose. Our main discovery is that the best-known (asymptotically optimal) resilient data structures have very large space overheads. More precisely, most of the space used by these data structures is not due to key storage. This might not be acceptable in practice since resilient data structures are meant for applications where a huge amount of data (often of the order of terabytes) has to be stored. Exploiting techniques developed in the context of resilient (static) sorting and searching, in combination with some new ideas, we designed and engineered an alternative implementation which, while still guaranteeing optimal asymptotic time and space bounds, performs much better in terms of memory without compromising the time efficiency.

## 1 Introduction

Memories in modern computing platforms are not always fully reliable, and sometimes the content of a memory word may be corrupted. This may depend on manufacturing defects, power failures, or environmental conditions such as cosmic radiation and alpha particles [4]. The classical way to deal with memory faults is via error detection and correction mechanisms, such as redundancy, Hamming codes, etc. These traditional approaches imply non-negligible costs in terms of time and money.

In this research we focus on the faulty-RAM model introduced in [3]. In this model, an adaptive adversary can corrupt any memory word, at any time, by overwriting its value. Corrupted values cannot be (directly) distinguished from correct ones. An upper bound  $\delta$  is given on the total number of memory faults that can occur throughout the execution of an algorithm or during the lifetime of a data structure. However, we can exploit  $O(1)$  safe memory words, whose content never gets corrupted.

A natural approach to the design of algorithms and data structures in the faulty-RAM model is data replication. Informally, a resilient variable consists of  $(2\delta + 1)$  copies  $x_1, x_2, \dots, x_{2\delta+1}$  of a standard variable. The value of a resilient variable is given by the majority of its copies (which can be computed in linear time and constant space). Observe that the value of  $x$  is reliable, since the adversary cannot corrupt the majority of its copies. We say that an algorithm or data structure is resilient to memory faults if, despite the corruption of some memory location during its lifetime, it is nevertheless able to operate correctly (at least) on the set of uncorrupted values.

### 1.1 Resilient Dictionaries

More recently, the problem of implementing resilient data structures has been addressed. A resilient dictionary is a dictionary where the insert and delete operations are defined as usual, while the search operation must be resilient as described before. In [2], Finocchi et al. present a resilient

dictionary using  $O(\log n + \delta^2)$  amortized time per operation. In [1], Brodal et al. present a simple randomized algorithm Rand achieving optimal  $O(\log n + \delta)$  time per operation. Using an alternative, more sophisticated approach, they also obtain a deterministic resilient dictionary Det with the same asymptotic performances. For all the mentioned implementations, the space usage is  $O(n)$ , which is optimal. However the constant hidden in the latter bound is not negligible in practice.

In this research we focus our attention on resilient dictionaries. We perform an experimental evaluation of the optimal dictionaries Det and Rand, together with an improved implementation RandMem developed by ourselves. In order to evaluate the drawbacks and benefits of resilient implementations, we also consider a standard (non-resilient) implementation of a search tree. In particular, we implemented an AVL binary search tree called Avl, in order to make a more direct comparison with Rand (which builds upon the same data structure).

In order to test different data structures, we developed a testbed to simulate the faulty-RAM model. We model the data structure and the adversary as two separate parallel threads. The data structure thread simply runs the data structure considered on the input sequence of operations. The adversary thread is responsible for injecting  $\alpha \leq \delta$  faults during the lifetime of the data structure. In order to inject one fault, the adversary selects one unsafe memory word (among the ones used by the data structure) uniformly at random, and overwrites it with a random value.

We performed experiments both on random inputs and on real-world inputs. In random inputs, the instances consist of a sequence of random operations. A random insert simply inserts a random value in a given range  $[e, r]$  (the actual range is not really relevant). In a random search we search for a key  $\kappa$ , where, with probability  $1/2$ ,  $\kappa$  is chosen uniformly at random among the keys currently in the dictionary, and otherwise is set to a random value in  $[e, r]$ . In a random delete, we delete a random key  $\kappa$ , where  $\kappa$  is generated as in the case of the random search. We also performed experiments with non-random instances.

## 2 Results

The main findings of our experimental study are the following:

- Avl is much faster than the optimal resilient dictionaries Det and Rand. This suggests that resilient implementations should be used only when the risk of memory faults is concrete.
- The running time of Rand and Det is rather sensitive to the upper bound  $\delta$  on the number of memory faults, but not much to the actual number of faults  $\alpha$ . This is an indication that  $\delta$  has to be fixed very carefully: underestimating  $\delta$  compromises the resiliency while overestimating it might increase the running time.
- The space usage of Det and Rand is not substantially affected by  $\delta$ . For both data structures, the space usage is much larger than the lower bound given by the total size of the keys. This might not be acceptable in practical applications, where keys occupy terabytes of space. We also observed that Avl, which is even non-resilient, has a similar drawback.
- Our improved dictionary RandMem uses an amount of memory closer to the lower bound given by key storage, and has a running time close to (sometimes better than) Rand and Det. This makes RandMem probably the data structure of choice for practical applications.

## References

- 1 G. S. Brodal, R. Fagerberg, I. Finocchi, F. Grandoni, G. F. Italiano, A. G. Jorgensen, G. Moruz e T. Molhave. Optimal Resilient Dynamic Dictionaries. Proc. 15th Annual European Symposium on Algorithms (ESA'07), 347–358, 2007.
- 2 I. Finocchi, F. Grandoni, and G. F. Italiano. Resilient search trees. Proc. 18th ACM-SIAM Symposium on Discrete Algorithms (SODA'07), 547–553, 2007.
- 3 I. Finocchi, F. Grandoni, and G. F. Italiano. Designing reliable algorithms in unreliable memories. Computer Science Review, 1(2), 77–87, 2007.
- 4 S. Hamdioui, Z. Al-Ars, J. V. de Goor, and M. Rodgers. Dynamic faults in random-access-memories: Concept, faults models and tests. Journal of Electronic Testing: Theory and Applications, 19:195–205, 2003.

# RANDOMLY STOPPED NONLINEAR FRACTIONAL BIRTH PROCESSES

FEDERICO POLITO & ENZO ORSINGHER

We present and analyse the nonlinear classical pure birth process  $\mathcal{N}(t)$ ,  $t > 0$ , and the fractional pure birth process  $\mathcal{N}^\nu(t)$ ,  $t > 0$ , subordinated to various random times, namely the first-passage time  $T_t$  of the standard Brownian motion  $B(t)$ ,  $t > 0$ , the  $\alpha$ -stable subordinator  $\mathcal{S}^\alpha(t)$ ,  $\alpha \in (0, 1)$ , and others. For all of them we derive the state probability distribution  $\hat{p}_k(t)$ ,  $k \geq 1$  and, in some cases, we also present the corresponding governing differential equation.

We also highlight interesting interpretations for both the subordinated classical birth process  $\hat{\mathcal{N}}(t)$ ,  $t > 0$ , and its fractional counterpart  $\hat{\mathcal{N}}^\nu(t)$ ,  $t > 0$  in terms of classical birth processes with random rates evaluated on a stretched or squashed time scale.

Various types of compositions of the fractional pure birth process  $\mathcal{N}^\nu(t)$  have been examined in the last part of the paper. In particular, the processes  $\mathcal{N}^\nu(T_t)$ ,  $\mathcal{N}^\nu(\mathcal{S}^\alpha(t))$ ,  $\mathcal{N}^\nu(T_{2\nu}(t))$ , have been analysed, where  $T_{2\nu}(t)$ ,  $t > 0$ , is a process related to fractional diffusion equations. Also the related process  $\mathcal{N}(\mathcal{S}^\alpha(T_{2\nu}(t)))$  is investigated and compared with  $\mathcal{N}(T_{2\nu}(\mathcal{S}^\alpha(t))) = \mathcal{N}^\nu(\mathcal{S}^\alpha(t))$ . As a byproduct of our analysis, some formulae relating Mittag-Leffler functions are obtained.

## REFERENCES

- [1] Orsingher, E., Polito, F., *Fractional Pure Birth Processes*. Bernoulli. Vol. 16 (3), 858-881, August 2010.
- [2] Orsingher, E., Polito, F., *On a Fractional Linear Birth-Death Processes*. To appear in Bernoulli. 2011.
- [3] Orsingher, E., Polito, F., Sakhno, L., *Fractional Non-Linear, Linear and Sublinear Death Processes*. Journal of Statistical Physics, Vol. 141 (1), 68-93, October 2010.
- [4] Orsingher, E., Polito, F., *Randomly Stopped Nonlinear Fractional Birth Processes*. Submitted.

DIPARTIMENTO DI SCIENZE STATISTICHE, SAPIENZA UNIVERSITÀ DI ROMA  
*E-mail address:* federico.polito@uniroma1.it

## LA DISTRIBUZIONE DEI CONTRATTI DI ASSICURAZIONE ATTRAVERSO IL CANALE BANCARIO

Claudio Russo

La più recente dottrina, superata la tradizionale dicotomia assicurazioni vita – assicurazioni danni, è oramai orientata a rivenire il tratto unificante del contratto di assicurazione nel fatto che questo è, per sua natura, un contratto “naturalmente” intermediato, nel senso che l’interposizione di un intermediario (sia esso un agente, un broker o uno dei cd. canali alternativi) costituisce un tratto caratterizzante la sua stessa definizione su un piano normativo.

Il profondo impatto di tale conclusione – anch’essa da verificare nei suoi presupposti concettuali – impone all’interprete di verificare non solo il *trend* della recente produzione normativa, sia essa primaria o di vigilanza, ma il livello di tutela cui essa tende, con particolare riferimento ai contratti aventi un notevole impatto sociale (e comunque di massa: si pensi all’assicurazione r.c.a.).

Nella divisata prospettiva, assumono particolare rilievo i Regolamenti Isvap nn. 5 e 35, il primo teso a dettare le regole della distribuzione – ed avente l’indubbio merito di avere colmato un evidente vuoto normativo, fonte di evidenti diseconomie oltre che di una inadeguata tutela della clientela - ed il secondo a dare un contenuto concreto alla regola di trasparenza, nella convinzione che la clientela non è in grado di percepire, quanto meno immediatamente, il contenuto negoziale del contratto ed i profili economici ad esso connessi, specie per quanto concerne i costi correlati all’intermediazione.

La regola di trasparenza in ambito assicurativo, dunque, assume un significato tecnico, in quanto tale ulteriore rispetto a quanto presupposto nelle norme generali in materia di contratti con i consumatori, tendendosi ad evitare catene distributive troppo “lunghe” – comportanti ad avviso dell’Istituto un inutile incremento del costo finale della copertura – ed a rendere consapevoli gli assicurandi degli elementi costitutivi del premio di tariffa applicato.

In particolare, l’Isvap aveva dettato nel Reg. n. 35 una regola generale, vietando all’intermediario di potere essere nel contempo anche beneficiario della somma assicurata (l’art. 52).

E’ evidente come tale divieto, pur adottato senza particolari distinguo tra i vari intermediari, andasse tuttavia ad incidere sulla principale modalità distributiva dei prodotti assicurativi (danni e vita) attraverso il canale bancario. Costituisce infatti un dato di comune esperienza che la banca erogante il credito, non solo pretenda di essere adeguatamente garantita nel caso di eventi sfavorevoli incidenti della sfera giuridica del soggetto finanziato (decesso, IPT superiore ad una determinata soglia, perdita di occupazione ecc.), ma sia essa stessa ad offrire tali coperture sulla base di una o più polizze collettive previamente sottoscritte con imprese di assicurazione. Attività, quest’ultima, a fronte della quale viene remunerata da parte delle imprese in ragione della raccolta dei contratti assicurativi presso la sua clientela.

La norma è stata annullata dal TAR per vizi di procedura e non per ragioni attinenti il merito, ed è stata quindi recentemente riproposta in pubblica consultazione da parte dell’Organo di vigilanza.

Scopo della ricerca è di verificare la coerenza di tale intervento rispetto alla *ratio* profonda del sistema, e ciò sia alla stregua della pregressa normativa primaria e secondaria sia avendo riguardo agli interventi del legislatore comunitario.

Prodromica a qualsiasi conclusione, quale che ne sia il segno, è poi la scomposizione della fattispecie nei suoi elementi costitutivi, nel tentativo di evidenziarne i tratti informanti e gli eventuali scollegamenti rispetto al quadro di riferimento.

## Longevity Risk and Financial Market: some issues

*Fabio Baione*  
*Università del Sannio, Benevento*  
[fabio.baione@studioacra.it](mailto:fabio.baione@studioacra.it)

*Paolo De Angelis*  
*Università Sapienza, Roma*  
[paolo.deangelis@uniroma1.it](mailto:paolo.deangelis@uniroma1.it)

*Andrea Fortunati*  
*Università Sapienza, Roma*  
[andrea.fortunati@uniroma1.it](mailto:andrea.fortunati@uniroma1.it)

*Agostino Tripodi*  
*Università Sapienza, Roma*  
[agostino.tripodi@uniroma1.it](mailto:agostino.tripodi@uniroma1.it)

**Abstract:** Longevity risk (LR) is undoubtedly a much discussed actuarial subject at international level; it implicates uncertainty of mortality rates in the long term (i.e. unanticipated inter-temporal variations) and the consequent effects on long term survival probabilities. Public and private annuities providers (as pension funds and insurance companies) are involved in hedging the LR; a classical decision procedure can be structured to solve an asset-liability mismatching problem, introducing financial tools with payoff linked to longevity or mortality indexes.

## *Research patterns, trends and new challenges in cluster analysis<sup>(\*)</sup>*

Maurizio Vichi  
Sapienza University of Rome

Methodologies for cluster analysis are among the most well-known and appreciated statistical techniques of multivariate analysis. In the last twenty years they have been increasingly applied in new disciplines and frequently almost reinvented in many area of research such as computer science, engineering, bioinformatics and in specific fields including machine learning, data mining and pattern recognition. In this presentation we show new *statistical* research patterns and trends in methodologies for clustering. The illustrated methods have in common the statistical approach of formulating a mathematical model, estimating its parameters and finally fitting the model to data.

The presentation is divided in three parts: single clustering of a set of units, multi-partitioning of a set of units and a set of variables and clustering of longitudinal multivariate observations.

### *Model-Based partitioning and hierarchical clustering*

The cluster analysis problem of partitioning or hierarchical clustering a set of units from dissimilarity data is here handled with the statistical model-based approach of fitting the “closest” *classification matrix* to the observed dissimilarities. A classification matrix represents a clustering model expressed in terms of dissimilarities.

Three models for partitioning a set of units from dissimilarity data, are illustrated and their estimation -via least-squares- is given together with new fast coordinate descent algorithms. Following the same statistical fitting approach a new model for hierarchical clustering from dissimilarity data is also illustrated.

### *Bi-partitioning, multi-partitioning, clustering and disjoint principal component analysis*

New methodologies for two-mode (units and variables) multi-partitioning of two way data are presented. In particular, by reanalyzing the *double k-means*, that identifies a unique partition for each mode of the data, a relevant extension is discussed which allows to specify more partitions of one mode, conditionally to the partition of the other one. The performance of such *generalized double k-means* has been tested by both a simulation study and an application to gene microarray data. Clustering and disjoint principal component allows to identify a partition of the units and a partition of the variables together with a principal component for each class of the partition of the variables. This technique can be seen as a special case of the bi-partitioning.

### *Clustering longitudinal multivariate observations.*

Longitudinal multivariate data involve repeated observations of different features of the same statistical units over a period of time. The aim is to study the developmental trends of the units across at least a part of their life span.

The dynamic evolution of the partitions of units along time is in this presentation studied in an unsupervised clustering context using a model based clustering approach. A clustering together with a vector autoregression VAR( $P$ ) model -where  $P$  is the lag length of the VAR- are combined into a new technique that identifies an homogeneous partition in  $G$  classes for

---

<sup>(\*)</sup> *This is part of the keynote lecture to be presented at the meeting of the Netherlands Society for Statistics, 3<sup>rd</sup> of March 2011*

each time  $t$  and the autoregressive dynamic evolution of the clusters. The proposed clustering/VAR model can be used also to forecast a partition at time  $T+1$ . The parameters of the model are estimated both in a least-squares and maximum likelihood framework and efficient recursive algorithms are given. A simulation study together with some applications of the proposed methodologies are shown to appreciate performances of the models and the quality of its estimates.

In the final part of the presentation, similarities between trajectories describing histories of units are studied. Trend, velocity and acceleration are three characteristics of trajectories considered to assess pairwise dissimilarities between trajectories. The Tucker model for three-way data, modified for clustering units together with a dimensional reduction of the observed variables, is estimated in the metric space specified by trend, velocity and acceleration. An application is given to show the performances of the methodology.

#### References

- Martella F., Alfò M., Vichi M. (2010). Hierarchical mixture models for biclustering in microarray, *Statistical Modelling*. To Appear.
- Rocci R., Vichi M. (2008). Two-mode Multi-partitioning. *Computational Statistics & Data Analysis*, vol. 52, pp. 1984-2003 ISSN: 0167-9473.
- Vicari D., Vichi M. (2011). On Multivariate Linear Regression for Heterogeneous Data, Submitted.
- Vichi M., Saporta G. (2009). Clustering and Disjoint Principal Component Analysis. *Computational Statistics & Data Analysis* vol. 53; p. 3194-3208, ISSN: 0167-9473, doi: 10.1016/j.csda.2008.05.028
- Vichi M. (2008) Fitting Semiparametric Clustering Models to Dissimilarity Data, *Advances in Data Analysis and Classification*, vol. 2, 2, 121-161.
- Vichi M. (2011) Fitting Hierarchical Clustering Models to Dissimilarity Data, Submitted.