

**Relazione scientifica di conferma in ruolo di professore di seconda fascia
di Eva Riccomagno, dicembre 2006-novembre 2009**

Pubblicazioni 2007-2009

Riviste internazionali

- a. *A brief history of algebraic statistics* (opening lecture of the 8th German Open Conference of Probability and Statistics). *Metrika* 69, 397–418, 2009.*
- b. (with Y Berstein, J Lee, H Maruri-Aguilar, S Onn, R Weismantel, H P Wynn) *Nonlinear Matroid Optimization and Experimental Design*. *SIAM Journal on Discrete Mathematics* 22:3, 901–919, 2008.
- c. (with H Maruri-Aguilar and R Notari) *On the description and identifiability analysis of mixture designs*. *Statistica Sinica*, 17:4, 1417-1440, 2007.*
- d. (with R Notari and M-P Rogantin) *Two polynomial representations of experimental design*. *Journal of Statistical Theory and Practice*, 1:3-4, 329-346, 2007.

Volumi collettanei

- e. (with P Gibilisco, M-P Rogantin and H P Wynn eds.) *Algebraic and geometric methods in statistics*. Cambridge University Press, 2009, 382 pp.*
 - e1. (with P Gibilisco, M-P Rogantin and H P Wynn eds.) Preface of *Algebraic and geometric methods in statistics* pages xiii-ix.
 - e2. (with P Gibilisco, M-P Rogantin and H P Wynn eds.) *Algebraic and geometric methods in statistics*. In *Algebraic and geometric methods in statistics*, Chapter 1 pages 1–23.
 - e3. (with R Notari) *Replicated measurements and algebraic statistics*. In *Algebraic and geometric methods in statistics*, Chapter 11 pages 187–201.
- f1. (with G Pistone and M-P Rogantin). *Methods in Algebraic Statistics for the Design of Experiments*. In *Search for Optimality in Design and Statistics: Algebraic and Dynamical System Methods* (L Pronzato and A A Zigljavsky eds.) Springer-Verlag 97–132, 2009.
- f2. (with J Q Smith) *The geometry of causal probability trees that are algebraically constrained*. In *Search for Optimality in Design and Statistics: Algebraic and Dynamical System Methods* (L Pronzato and A A Zigljavsky eds.) Springer-Verlag 133–154, 2009.
- g. (with H P Wynn). *An introduction to regression and errors in variables from an algebraic viewpoint*. In *Approximate Commutative Algebra* (L Robbiano and J Abbott eds.) Springer-Verlag 193–203, 2009.
- h. (with H Maruri-Aguilar) *A model selection algorithm for mixture experiments including process variables*. *Proceedings of Moda8* (J Lopez-Fidalgo, J Rodríguez-Díaz, B Torsney eds.) Springer 107–114, 2007.

Quaderni di dipartimento e altre pubblicazioni

- i. (with Y Berstein, H Maruri-Aguilar, S Onn, H P Wynn) *Minimal average degree aberration and the state polytope for experimental design* (under revision for *Annals of Institute of Statistical Mathematics*).*
- j. (with J Q Smith and P Thwaites) *Causal analysis with chain event graphs* (under revision for *Artificial Intelligence*).
- k. (with J Q Smith) *Algebraic Causality: Bayes nets and beyond*, arxiv.org/abs/0709.3377, CRiSM Paper No. 07-07, 2007.
- l. (with J Q Smith) *The causal manipulation of chain event graphs*, arxiv.org/abs/0709.3380, 2007.*

Le pubblicazioni sono peer-reviewed da due o più reviewer eccetto e1, e2, g, l. Gli autori sono in ordine alfabetico. Al 5 gennaio 2010 su MathSciNet sono apparse a, b, c, d, f1, f2, h.

Conferenze e workshops

Attività di organizzazione

- 2009 Organizzatrice con M Drton (Chicago University) e S Sullivant (North Carolina State University) del workshop *Algebraic Statistical Models* parte del Thematic Year on Algebraic Methods in Systems Biology and Statistics, SAMSI 2008/9, 14-17 gennaio, USA.
- 2008 a. Responsabile scientifico con P Gibilisco (Università di Roma “Tor Vergata”) del convegno *Mathematical explorations in contemporary Statistics*, 19-20 maggio, Sestri Levante.*
- b. Organizzatrice della sessione invitata *Algebraic Statistics* al 7th World Congress in Probability and Statistics, 14-19 luglio, Singapore.*
- 2007 Membro del comitato scientifico di ENBIS7, 24-26 settembre, Dortmund, De, www.enbis.org/.

Presentazioni

- 2009 (a) Roundtable discussant at WOGAS1, Workshop on Geometric and Algebraic Statistics, 13 luglio, Open University UK, stats-www.open.ac.uk/WOGAS.
- (b) Lecture course on *Algebraic Statistics* at the Second de Brún Workshop on Computational Algebra, Galway, Ireland, 29 giugno-11 luglio.*
- (c) 7th Workshop on Bayesian Nonparametrics, 21-25 giugno, Torino, IT. Poster *Causal analysis with chain event graphs*.
- (d) Giornata di Statistica Algebrica, 28 maggio, Torino. *Statistica Algebrica: alcune idee e spunti per la ricerca*.
- (e) Partecipante al workshop Subjective Bayes, 15-17 aprile, Warwick.
- (f) Algebraic Statistical Models Workshop, 15-17 gennaio, SAMSI, USA. Poster *Causal analysis with chain event graphs*, www.samsi.info/.
- 2008 (a) Invited speaker a CASTA2008, Computational Algebraic Statistics, Theories and Applications, 10-11 dicembre, Kyoto, Jp.. *Repeated measurements and algebraic statistics*.*
- (b) Invited speaker alla sessione invitata Recent developments in algebraic statistics in IASC 2008, International Association for Statistical Computing, 5-8 dicembre, Yokohama, Japan. *Between algebraic statistics and information geometry*.
- (c) Partecipante all’opening workshop del 2008/9 SAMSI Thematic Year on Algebraic Methods in Systems Biology and Statistics, 14-17 settembre, USA. *Between algebraic statistics and information geometry*.
- (d) Lezione di apertura dell’8th German Open Conference of Probability and Statistics in 3-7 marzo, Aquisgrana, De. *A brief history of algebraic statistics*.*
- 2007 (a) IMA Workshop su Applications in Biology, Dynamics, and Statistics parte dell’anno tematico in Applications of Algebraic Geometry, 5-9 marzo, Minnesota.
- (b) ESF Exploratory Workshop in Multivariate Interpolation, 16-19 ottobre, Sestri Levante. *Replicated measurements and algebraic statistics*.

Invitata a tutte le conferenze precedenti e finanziata per tutte eccetto 2007(b), 2009(a,c,d).

Affiliazioni e visite ad istituti scientifici

- Associate Fellow presso il Department of Statistics, The University of Warwick, UK.
- Associate Member di CRiSM, Center for Research in Statistical Methodology, Warwick, UK.
- 12-25 Gennaio 2009 Research Fellow al SAMSI, Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC, USA.

Supporto alla ricerca

- 2009 membro esperto per il “panel Mathematics” del Research Council Norvegese, Frinat program.
- 13-14 marzo 2008. Componente esperto della commissione d’esame di 9 tesi per il Dottorato di Ricerca in Matematica per le Scienze dell’Ingegneria, Politecnico di Torino.
- 2007 novembre. Membro della commissione d’esame per la tesi di dottorato *Observability and identifiability of non-linear systems with applications in biology* candidata M Anguelova, Göteborg, Svezia.
- 2007 Hugo Maruri-Aguilar discute la tesi “Methods from computational commutative algebra in design and analysis of experiments” per un Ph.D. in statistica sotto la mia direzione.
- Reviewer per Journal of Computational and Applied Mathematics, Journal of the Royal Society: Interface, Journal of Theoretical Biology, Bernoulli, Metrika, Journal of Statistical Theory and Practice. Inoltre per rivista italiana Metron, per Moda9 proceedings e per Mathematical Reviews.
- Membro delle seguenti associazioni ed organizzazioni scientifiche: INDAM-GNAMPA, RSS, BS, SIS, SIMAI, ISBA, ISM e ENBIS (chartered member, www.enbis.org).

Commenti

In questi tre anni mi sono impegnata nella divulgazione e survey di temi legati alla statistica algebrica (AS), sono stata uno dei principali fattori dell'avvicinamento della geometria dell'informazione e della statistica algebrica (AS+IG) e ho pubblicato in collaborazione risultati originali su analisi di dati composizionali (mixture), misure ripetute (replicates), un nuovo concetto di robustezza per piani sperimentali (algebraic aberration), ho introdotto modelli e tecniche di analisi per studiare relazioni di causalità in dati categorici (causality). Nella tabella successiva, dove '-x' indica l'anno '200x', sono grossolanamente suddivisi gli elementi degli elenchi precedenti. Un * negli elenchi indica attività che considero particolarmente significative.

	dissemination	AS	AS+IG	mixture	aberration	replicates	causality	altri
Publ.	a. f1. g.		e. e1. e2.	c. h.	b. i.	e3.	f2. j. k. l.	d.
Organ.	-9 -8b.		-8a.					-7
Talks	-9(b,d)		-9(a) -8(b,c)			-8(a) -7(b)	-9(c,e,f)	-8(d)

Statistica algebrica

Recentemente si è presentata una nuova prospettiva per lo studio di modelli statistici e per l'analisi statistica, grazie allo sviluppo ed implementazione di software per il calcolo con polinomi multivariati. Molti modelli statistici hanno una struttura polinomiale in opportune parametrizzazioni; molte procedure inferenziali richiedono la soluzione di equazioni polinomiali, tipicamente quelle basate sulla verosimiglianza; un piano sperimentale ed il supporto di un vettore aleatorio discreto possono essere rappresentati come l'insieme delle soluzioni di un sistema polinomiale; negli ultimi tre anni generalizzazioni a vettori aleatori continui sono state studiate.

Partendo da queste osservazioni, si è sviluppata una nuova area interdisciplinare di ricerca teorica ed applicata chiamata Statistica Algebrica che coinvolge tra gli altri statistici, matematici, biologi e che sta assumendo un ruolo di rilevanza internazionale nella statistica matematica e in algebra computazionale. Ciò è testimoniato da almeno cinque libri e due numeri speciali di riviste scientifiche dedicati alla statistica algebrica, dall'anno tematico 2008 al SAMSI e da oltre una ventina di conferenze.

Il mio contributo per il 2007-09 è illustrato, tra l'altro, dall'invito ad organizzare una sessione per uno dei più importanti convegni di statistica matematica e probabilità (2008b.), ad inaugurare una delle più seguite conferenze nazionali in Europa di statistica matematica e analisi stocastica con oltre 500 delegati (2008(d), a., gocps2008.rwth-aachen.de/), a presentare un corso intensivo al De Brún center for computational algebra (2009(b), hamilton.nuigalway.ie/DeBrunCentre/).

Statistica algebrica e geometria dell'informazione

Le interazioni tra due teorie tradizionali della matematica, la geometria algebrica e la geometria differenziale, con la statistica matematica e la probabilità hanno prodotto da un lato la statistica algebrica e dall'altro la geometria dell'informazione (Amari, 2000). Sulla statistica algebrica abbiamo scritto sopra, circa la geometria dell'informazione citiamo solo che l'informazione di Fisher, oggetto fondamentale nell'inferenza statistica, è una metrica Riemanniana su un modello statistico e che la ricerca attuale si concentra sullo studio della geometria indotta da un modello di classe esponenziale sui suoi sottomodelli. La realizzazione che importanti nozioni di statistica e probabilità hanno una natura algebrica e/o geometrica ha stimolato la ricerca a livello fondazionale. Ma solo recentemente è stata evidenziata una relazione diretta tra queste due branche della statistica matematica, soprattutto grazie a due seminari di G. Pistone datati 2005 e 2007 (cf. 2007(a)).

Il volume e., peer-reviewed eccetto e1., e2., è stato ideato con l'intento di mettere in contatto le due comunità di statistica algebrica e geometria dell'informazione. Il successo di questo progetto, in parte diretto ed in parte indiretto, è intuibile dall'istituzione di WOGAS1 (cf. 2009(a)) e dalla programmazione di WOGAS2 per aprile 2010; dall'accoglienza internazionale del convegno 2008a; e dall'attività di alcuni dottorandi in Germania, Spagna, Milano, che stanno sviluppando risultati nello studio di invarianti di modelli statistici, di famiglie esponenziali, con applicazioni in filogenetica ed intelligenza artificiale. Più avventurosamente, ho iniziato a considerare possibili sviluppi in analisi stocastica.

Statistica algebrica e piani sperimentali

Gli articoli c. e h. presentano una metodologia che risponde a quesiti irrisolti dagli anni '60 sui piani sperimentali per dati composizionali (si vedano Sheffé 1958, 1963 per la formulazione dei quesiti e Cornell 2002 per una revisione dei risultati, parziali, ottenuti nel frattempo). Questa tipologia di dato ha forte valenza applicativa ed è tipica di discipline dove è importante la "composizione" e.g. minerali in una roccia. Gli articoli b. e i. sviluppano, da un punto di vista algebrico e statistico rispettivamente, un indice che esprime la potenzialità di un piano sperimentale nell'identificare una famiglia di modelli di regressione polinomiale: più alto è l'indice più tanti sono i modelli polinomiali di regressione identificabili e quindi maggiore è la salvaguardia contro un modello inadeguato. Questi lavori sono parte o sviluppi della tesi di Ph.D. di Hugo Maruri-Aguilar, dottoratosi sotto la mia supervisione. Presentazioni su questi temi a convegni e workshop internazionali sono state fatte principalmente da Hugo, ora lecturer alla Queen Mary, University of London.

Nell'articolo d. sono sviluppati algoritmi per passare tra le due rappresentazioni note in statistica algebrica di un piano sperimentale, specificamente funzione indicatrice e spazio quoziente. I vantaggi di questo passaggio sono illustrati nel paragrafo 5 di f1., che altrimenti presenta lo stato dell'arte al 2008 della statistica algebrica per il disegno degli esperimenti. In e3. è generalizzato l'approccio algebrico all'analisi di piani sperimentali con esperimenti ripetuti ed un breve paragrafo, che necessita di essere sviluppato, suggerisce che sviluppi allo studio della varianza e di momenti superiori necessitano strumenti di geometria differenziale.

La rilevanza di questa ricerca, e più generalmente della statistica algebrica, per le applicazioni è che fornisce metodologie per analizzare e pianificare esperimenti non standard, oltre a fornire strumenti per comprendere appieno modelli e risultati inferenziali.

Causalità: chain event graph

Fondamentale in statistica, e non solo, è una corretta modellizzazione ed implementazione del concetto di causalità, da non confondersi con associazione né indipendenza. In letteratura statistica sia teorica che applicata sono principalmente considerate tre teorie matematico-statistiche per la causalità: equazioni strutturali e modelli grafici (J. Pearl, 2000), potential outcomes e counterfactuals (D. Rubin, 2005 per un survey), affiancati più recentemente da quello di G. Shafer (2003), che affiancano la più classica Granger causality per serie temporali.

Una mia ricerca precedente al 2007 in collaborazione con J Q Smith (Warwick University, UK) ha portato al riconoscimento della natura algebrica dei modelli causali alla Pearl per un numero finito di variabili aleatorie discrete. Questi ammettono una parametrizzazione *implicita*, che potremmo definire 'torica', tramite un insieme di uguaglianze e disuguaglianze tra polinomi nelle probabilità congiunte. Questa interpretazione algebrica ha permesso la generalizzazione a situazioni non standard e alla definizione di nuove classi di modelli casuali e ha aperto il campo agli strumenti dell'algebra computazionale e della geometria (cf. f2., k.). La vasta generalità di questi modelli è limitata nelle applicazioni pratiche dalla mancanza di algoritmi veloci per calcoli di propagazione ed identificabilità.

Un'altra direzione di ricerca è stata ispirata dalla modellizzazione della nozione di causalità tramite alberi di probabilità, tipica di Shafer, in cui i nodi dell'albero rappresentano non necessariamente delle variabili aleatorie, ma piuttosto degli stati di variabili aleatorie o più in generale eventi di un processo. Con Jim abbiamo definito un modello, detto chain event graph, che si presenta come un albero orientato sui cui nodi sono definite due relazioni di equivalenza. Questo modello si basa su una parametrizzazione *esplicita* data dalle probabilità di transizione tra nodi. Anche questo modello ammette una rappresentazione tramite ideali di polinomi e disuguaglianze tra polinomi. Su questo modello abbiamo definito una nozione di causalità come intervento, cioè alla Pearl, studiato un semplice algoritmo di propagazione sotto campionamento multinomiale e considerato i rudimenti della questione dell'identificabilità (cf. l.). Successivamente con P Thwaites (Warwick University), abbiamo sviluppato dei teoremi ed algoritmi per l'identificazione di una "causa" (una certa funzione dei parametri) in termini di "variabili manifeste" (altre funzioni dei parametri) (cf. j.) che rendono questi modelli efficienti in alcune applicazioni.