

# Dipartimento di Matematica - Università degli Studi di Genova

**1. Linea di ricerca:** Statistica matematica: aspetti teorici, algebrici e applicazioni / SECS-S/01

## 2. Componenti

Eva Riccomagno / PA / SECS-S/01 /  
Maria Piera Rogantin / PA / SECS-S/01 /  
Claudia Fassino / Ric / MAT08 /

## 3. Collaborazioni principali

Hugo Maruri-Aguilar	Mathematical Sciences	Queen Mary, University of London
Bernd Sturmfels	Mathematics	Berkeley
Sonja Kuhnt	Mathematische Statistik und industrielle	Technische Universität Dortmund
Nikolaus Rudak	Mathematische Statistik und industrielle	Technische Universität Dortmund
Sofia Massa	Statistics	University of Oxford
James Q. Smith	Statistics	University of Warwick
Manuele Leonelli	Statistics	University of Warwick
Henry P. Wynn	Statistics	London School of Economics
Roberto Fontana	Matematica	Politecnico di Torino
Giovanni Pistone	de Castro Statistics Initiative	Collegio Carlo Alberto
Fabio Rapallo	DISIT	Università del Piemonte Orientale
Massimo Caccia	ISSIA	CNR Genova
Fabrizio Malfanti	Intelligrate	Genova

## 4. Temi di ricerca

### Tema 1: Statistica algebrica e applicazioni

Sottotema 1: Formule di cubatura (con Fassino e Pistone). Calcolo di valori attesi con metodi algebrici. Risultati ottenuti: algoritmo generale, caratterizzazione di formule di cubatura esatte. Ricerca attuale: stima dell'errore, applicazione al chaos polinomiale.

Sottotema 2: Analisi e pianificazione degli esperimenti. Risultati ottenuti tramite la funzione indicatrice (con Rapallo, Fontana: teoria e software per due fattori; ricerca attuale: generalizzare a più fattori tramite l'utilizzo di basi di Markov, di Gröbner universale e di Graver. Risultati ottenuti per modellazione (con Rudak e Kuhnt): teoria completa sviluppata; ricerca attuale: applicaziozne a modelli di regressione a catena.

Sottotema 3: Modelli torici per catene di Markov reversibili (con Pistone). La teoria dei modelli torici è abbastanza sviluppata in letteratura. La ricerca riguarda la nuova applicazione alle catene con preferenze personalizzate.

Sottotema 4: Disuguaglianze Estensione a variabili discrete dei risultati ottenuti per modelli filogenetici nel caso binario (con Piotr Zwiernik, borsa Marie Curie outgoing con Sturmfels referente a Berkeley) e caratterizzazione tramite invarianti polinomiali di funzioni di utilità multi-attributo con struttura grafica (con Leonelli, dottorando, e Smith).

Sottotema 5: Combinazioni markoviane di modelli grafici d'indipendenza (con Massa). In fase di completamento: modello Gaussiano. Da sviluppare: variabili aleatorie discrete.

Sottotema 6: Topologia algebrica per raster-plot (con Virginia Pirino / dottoranda, Sergio Martinoia / PA e Paolo Massorbio / Assegnista, DIBRIS). Costruzione ed interpretazione statistica di complessi simpliciali, numeri di Betti e ideali polinomiali.

### Tema 2: Definizione di metodologie sperimentali

Sottotema 1: Fraud detection in online banking. Sviluppo e validazione di metodi statistico-matematici e di data mining (con Malfanti e Delio Panaro / assegnista DIMA).

Sottotema 2: Pianificazione ottima degli esperimenti in veicoli marini senza equipaggio (con Caccia e Saggini, dottoranda DIMA). Ambito applicativo

Sottotema 3: v. Sottotema 5 del tema precedente.

## 5. Pubblicazioni principali

1. G. Pistone, E. Riccomagno, H. P. Wynn (2001). Algebraic Statistics. CRC Press.
2. Gibilisco, Riccomagno, Rogantin, Wynn (eds, 2009). Algebraic and geometric methods in statistics, Cambridge Univ. Press, Cambridge.

---

3. Pistone, Rogantin (2013). The algebra of reversible Markov chains. Ann Inst Stat Math, 65:26993.
4. Fontana, Rapallo, Rogantin (2012). Markov Bases for Sudoku Grids Statistical methods for the analysis of large data sets (Di Ciaccio, Coli, Anguol Ibanez Eds) Springer, 305–315.
5. Pistone, Rogantin (2008) Indicator function and complex coding for mixed fractional factorial designs. J Statist Plann Inference 138, 787–802

---

6. Fassino, Pistone, Riccomagno (2013). The algebra of interpolatory cubature formulae for generic nodes. Statistics and computing (in press).
7. Riccomagno, Smith and Thwaites (2010) Causal analysis with Chain Event Graphs Artificial Intelligence 174, 889–909

---

8. Riccomagno (2009) A brief history of algebraic statistics Metrika 69, 397-418

---

9. Riccomagno, Rogantin and Pistone (2009) Methods in Algebraic Statistics for the Design of Experiments In Search for Optimality in Design and Statistics (Pronzato and Zigljavsky Eds) 97–132

---

10. Riccomagno, R A Bates, R J Buck and H P Wynn (1996). Experimental Design and Observation for Large Systems (with discussion). Journal of the Royal Statistical Society, B, 58:1, 77–94.

## 6. Inquadramento della linea di ricerca in ambito internazionale

Sottotemi 1.1-1.5 hanno collaboratori internazionali, come indicato. Si inseriscono nell'ambito della statistica algebrica (AS). AS è un'area di ricerca interdisciplinare che coinvolge statistici, matematici, biologi e che ha assunto un ruolo di rilevanza internazionale, testimoniato da varie conferenze organizzate ogni anno principalmente in USA, Europa, Giappone; la fondazione recente di una rivista, vari numeri dedicati di riviste matematiche, statistiche, applicative; parecchie posizioni accademiche e di ricerca assegnate a dottorati con tesi di SA (e.g. Maruri-Aguilar, Ruriko Yoshida / Kentucky, Caroline Uhler / IST Austria).

I ricercatori genovesi oltre a sviluppare strumenti teorici caratterizzanti la disciplina hanno anche svolto ricerche applicate e spesso hanno implementato i loro risultati sulle piattaforme di calcolo più opportune.

Il nostro gruppo si è fortemente impegnato nello sviluppare l'interesse per AS anche tramite lo scambio di visite scientifiche, l'organizzazione e la partecipazione a convegni. L'attività di disseminazione si è concretizzata recentemente in un eterogeneo gruppo internazionale di giovani che, spesso sotto la codirezione nostra e di un relatore, proseguono e sviluppano la loro ricerca su specifici temi più applicativi rispetto a quelli classici della Statistica algebrica.

La definizione di combinazioni Markoviane di leggi di probabilità risale al 1993 (Dawid, Lauritzen, *Ann. Statist.*), ma solo recentemente la SA ha permesso di "automatizzare" le operazioni e rendere operative le combinazioni portando così alla nozione di meta-analisi strutturale. Questa è di particolare interesse in medicina e farmacologia dove si vuole combinare conclusioni ottenute da diversi esperimenti non solo quelle date da indici numeri (meta-analisi) ma anche i relativi modelli.

Il Sottotema 1.6 è parte della ricerca internazionale relativa allo studio della struttura funzionale del cervello. Le tecniche matematiche da noi adottate prendono spunto dalla Topological Data Analysis, e si basano su strumenti di topologia computazionale e di algebra commutativa computazionale. Autori della principale bibliografia matematica per il nostro lavoro sono Gunnar Carlsson e collaboratori, Carina Curto e collaboratori.

Relativamente al Sottotema 2.1 si osserva che è un problema di "big data", di cui tanto si parla, non solo nella comunità scientifica. Ci sono modelli, quasi tutti statistici e molte note informatiche, relative al problema analogo di "credit card fraud detection", alcuni di questi non sono trasponibili per l'online banking, per altri non si hanno stime di errori né di prima né seconda specie. Non abbiamo collaboratori internazionali su questo tema.

Il disegno o pianificazione degli esperimenti è un tema classico di Statistica Matematica che è trattato con strumenti di algebra polinomiale in SA nei temi di cui sopra. Il Sottotema 2.2 riguarda invece la pianificazione ottimale per parametri significativi in modelli alle equazioni differenziali e per grossi sistemi. Si adottano metodologie basate sull'ottimizzazione di un'opportuna funzione di una certa matrice di Fisher (v. pubblicazione principale 10). Riferimenti bibliografici principali sono legati all'ampio gruppo legato alla serie di Workshops on Model-Oriented Data Analysis and Optimum Design (MO-DA) in particolare Henry P Wynn che riunisce gli esperti mondiali su disegno ottimo degli esperimenti. L'applicazione di interesse e la provenienza dei fondi per la dottoranda E. Saggini (CNR-ISSIA) lo posizionano all'interno di un ampio progetto sulla definizione di un minimo insieme di esperimenti ripetibili (quali path following, target paths & initial conditions) per valutare la performance dei veicoli.

## 7. Finanziamenti

Miur, Rientro dei cervelli	Riccomagno PA	2005	2 anni (2005, 2006)
Provincia di Genova. Progetto DE.CI.DI.	Rogantin PA	terza fase 2006	1 anno
Provincia di Genova <sup>1</sup>	Rogantin PA	2006	1 anno
DAAD	Riccomagno PA	2012	2 mesi (6-7/2012)
FP7-PEOPLE-2011-IOF	Riccomagno PA	2013	3 anni (2013-14)
Fondazione CARIGE	Riccomagno PA	2013	1 anno (2013)