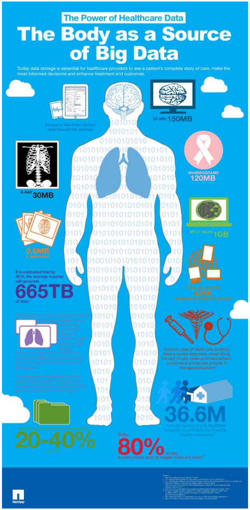
Presentazione ai dottorandi: The Cattolica reseach group

Raffaele Argiento
raffaele.argiento@unicatt.it

Bicocca
Milano, 24 Settembre 2020

Raffaele Argiento
raffaele.argiento@unicatt.it

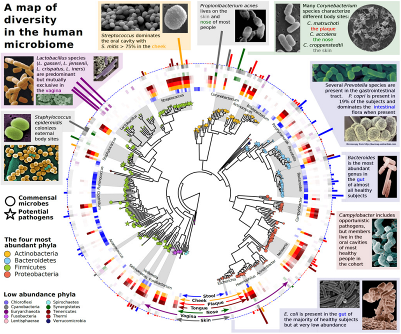
Bicocca
Milano, 24 Settembre 2020



Collaborations: Rice University & MD Anderson, TX USA – National Univerisy of Singapore (NUS) – College of Public Healt, National Taiwan University.

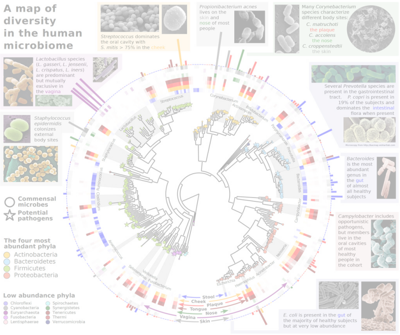
Collaborations: Rice University & MD Anderson, TX USA – National Univerisy of Singapore (NUS) – College of Public Healt, National Taiwan University.

Microbiome data: modelling the association between bacteria taxa e nutrients

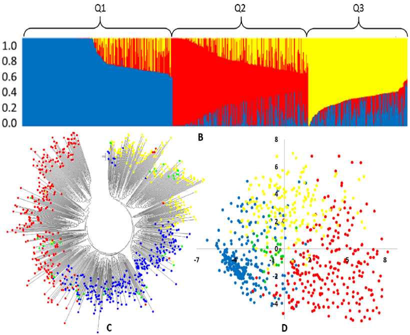


Collaborations: Rice University & MD Anderson, TX USA – National Univerisy of Singapore (NUS) – College of Public Healt, National Taiwan University.

Microbiome data: modelling the association between bacteria taxa e nutrients

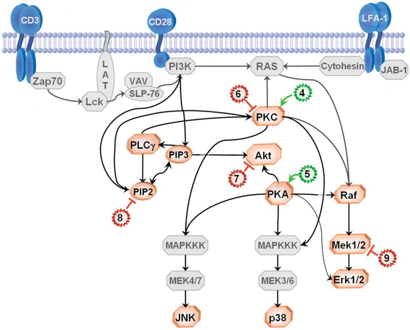


Population structure: Genetic Diversity (subpopulations)



Collaborations: University of Firenze – University of Modena e Reggio Emilia – National Technical University of Athens.

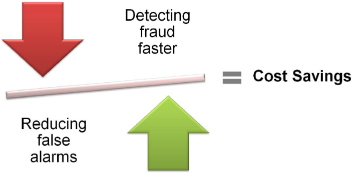
Protein data: Discover relationship between proteins, causal inference



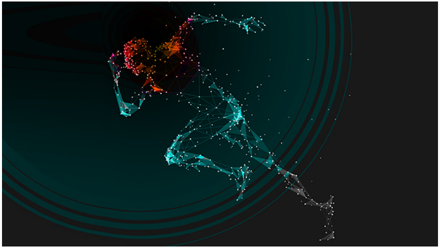
Collaborations: Università della Svizzera Italiana – Axa –

Protein data: Discover relationship between proteins, causal inference

Fraud detection: shorten the delay from the occurrence of the fraud to its detection;



Collaborations: Kent University (UK) – Università di Torino – Politecnico Milano – Victoria University, Melbourne (AUS)



Applications: Sport Data & Healthcare

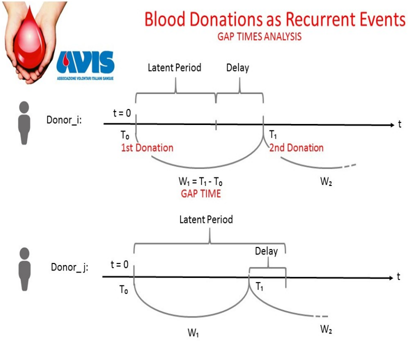
Collaborations: Kent University (UK) – Università di Torino – Politecnico Milano – Victoria University, Melbourne (AUS)

Sport analytics: performances predictions, doping detection

Collaborations: Kent University (UK) – Università di Torino – Politecnico Milano – Victoria University, Melbourne (AUS)

Sport analytics: performances predictions, doping detection

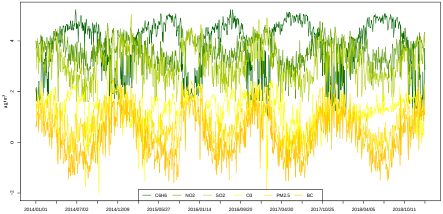
Avis data: Clustering donors for customized advertising



- Air pollution is a major global environmental risk to human health (WHO, 2018)
- We are simultaneously exposed to a complex mixture of air pollutants
- Moving toward a multi-pollutant approach to air quality

- Air pollution is a major global environmental risk to human health (WHO, 2018)
- We are simultaneously exposed to a complex mixture of air pollutants
- Moving toward a multi-pollutant approach to air quality

- A better understanding of the interactions between air pollutants is critical
- *Learning dependencies among multiple time series*



Hierarchical modelling

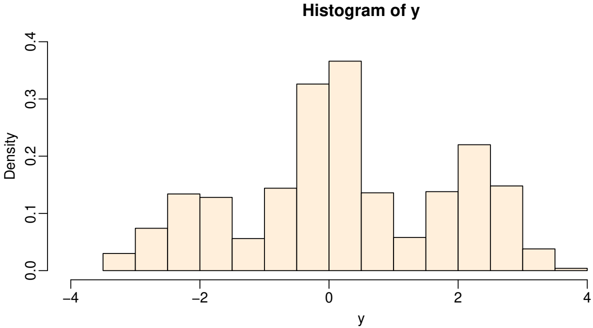
$$
\begin{aligned}
&Y_1, \ldots, Y_n | \theta_1, \ldots, \theta_n \overset{\text{ind.}}{\sim} f(y_i | \theta_i)\\
&\theta_1, \ldots, \theta_n | P \overset{\text{i.i.d.}}{\sim} P\\
&P(\cdot) \overset{d}{=} \textstyle\sum_{h=1}^{\infty} w_h \delta_{\tau_h}(\cdot)
\end{aligned}
$$

Hierarchical modelling

$$Y_1, \ldots, Y_n | \theta_1, \ldots, \theta_n \overset{\text{ind.}}{\sim} f(y_i|\theta_i)$$

$$\theta_1, \ldots, \theta_n | P \overset{\text{i.i.d.}}{\sim} P$$

$$P(\cdot) \overset{d}{=} \sum_{h=1}^{\infty} w_h \delta_{\tau_h}(\cdot) \sim Norm - CRM$$

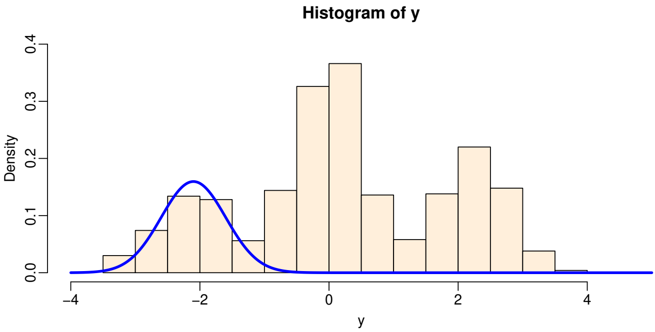


Hierarchical modelling

$$Y_1, \ldots, Y_n | \theta_1, \ldots, \theta_n \overset{\text{ind.}}{\sim} f(y_i|\theta_i)$$

$$\theta_1, \ldots, \theta_n | P \overset{\text{i.i.d.}}{\sim} P$$

$$P(\cdot) \overset{d}{=} \textstyle\sum_{h=1}^{\infty} w_h \delta_{\tau_h}(\cdot) \sim \color{red}{Norm - CRM}$$

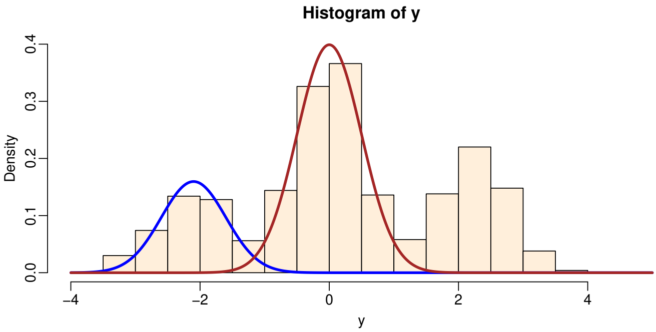


Hierarchical modelling

$$Y_1, \dots, Y_n | \theta_1, \dots, \theta_n \overset{\text{ind.}}{\sim} f(y_i|\theta_i)$$

$$\theta_1, \dots, \theta_n | P \overset{\text{i.i.d.}}{\sim} P$$

$$P(\cdot) \overset{d}{=} \sum_{h=1}^{\infty} w_h \delta_{\tau_h}(\cdot) \sim \mathit{Norm - CRM}$$

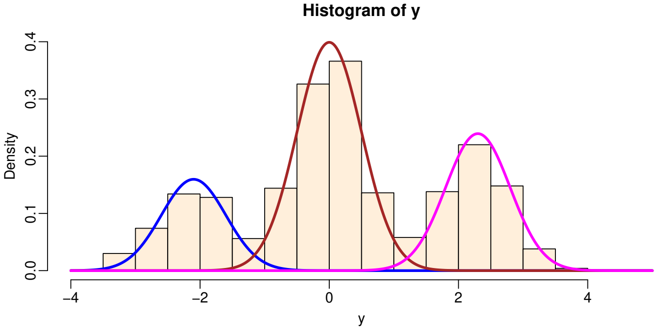


Hierarchical modelling

$$Y_1, \dots, Y_n | \theta_1, \dots, \theta_n \overset{\text{ind.}}{\sim} f(y_i | \theta_i)$$

$$\theta_1, \dots, \theta_n | P \overset{\text{i.i.d.}}{\sim} P$$

$$P(\cdot) \overset{d}{=} \sum_{h=1}^{\infty} w_h \delta_{\tau_h}(\cdot) \sim Norm - CRM$$

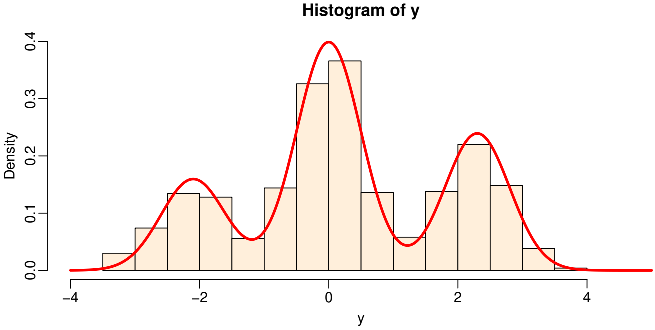


Hierarchical modelling

$$Y_1, \ldots, Y_n | \theta_1, \ldots, \theta_n \overset{\text{ind.}}{\sim} f(y_i|\theta_i)$$

$$\theta_1, \ldots, \theta_n | P \overset{\text{i.i.d.}}{\sim} P$$

$$P(\cdot) \overset{d}{=} \sum_{h=1}^{\infty} w_h \delta_{\tau_h}(\cdot) \sim Norm - CRM$$



Hierarchical modelling

$$Y_1, \ldots, Y_n | \theta_1, \ldots, \theta_n \overset{\text{ind.}}{\sim} f(y_i | \theta_i)$$

$$\theta_1, \ldots, \theta_n | P \overset{\text{i.i.d.}}{\sim} P$$

$$P(\cdot) \overset{d}{=} \sum_{h=1}^{\infty} w_h \delta_{\tau_h}(\cdot) \sim Norm - CRM$$

Ongoing works

(a) Dependent processes P_x, to include covariate information

(b) A general class of dependent models that encompasses many specific structures

(c) *Scalable algorithms* for fast inference

Selected Publications

- Argiento, R., Cremaschi, A. and Vannucci, M. (2019). "Hierarchical Normalized Completely Random Measures to Cluster Grouped Data", Journal of the American Statistical Association. Just accepted.
- Cremaschi, A., Argiento, R., Shoemaker, K., Peterson, C.B. and Vannucci M. (2019). "Hierarchical Normalized Completely Random Measures for Robust Graphical Modeling". Bayesian Analysis. Just accepted.
- Argiento R., Ruggiero, M. (2018). "Computational challenges and temporal dependence in Bayesian nonparametric models", Statistical Methods and Applications, Volume 27,

Graph theory

$$G = (V, E)$$

- finite set of **vertices** $V = \{1, \ldots, q\}$
- subset of **edges** $E \subseteq V \times V$

Nodes $\Leftrightarrow$ Random variables

Edges $\Leftrightarrow$ Probabilistic relationships

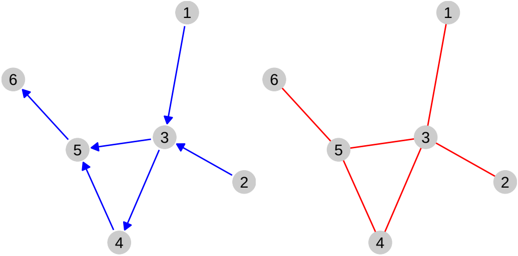


Figure: Directed (left) and undirected (right) graphs.

Graph theory

$$G = (V, E)$$

- finite set of **vertices** $V = \{1, \ldots, q\}$
- subset of **edges** $E \subseteq V \times V$

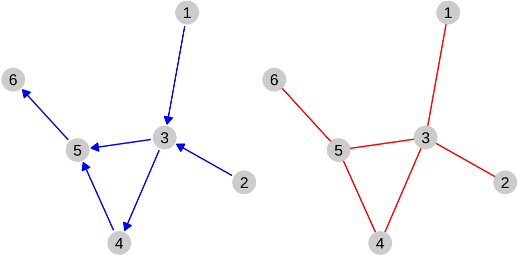


Figure: Directed (left) and undirected (right) graphs.

Nodes $\Leftrightarrow$ Random variables

Edges $\Leftrightarrow$ Probabilistic relationships

Graphical model

✓ Family of probability distributions for the q random variables which factorizes according to a given graph. ✓ *Conditional independencies* are read from the graph.

Graph theory

$$G = (V, E)$$

Figure: Directed (left) and undirected (right) graphs.

Ongoing works

(a) Objective Bayes Model Selection from Observational Data

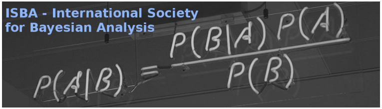
(b) Multiple Graphical Models

(c) Estimate Causal effects using Directed Graphical Models

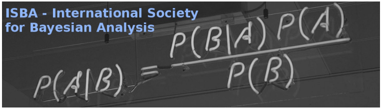
(c) Dependent graphs, spatio-temporal dependence to capture graphs relationships.

Pubblicazioni recenti

- Castelletti, F. Consonni, G., Della Vedova, M. L. & Peluso, S. (2018). "Learning Markov equivalence classes of Directed Acyclic Graphs: an Objective Bayes Approach." *Bayesian Analysis* **13**, 1231–1256.
- Castelletti, F. & Consonni, G. (2019). "Objective Bayes model selection of Gaussian interventional essential graphs for the identification of signaling pathways." *Annals of Applied Statistics*, in-press
- Paci, L. & Consonni, G. (2019). " Structural Learning of Contemporaneous Dependencies in Graphical VAR models", *Invited revision*



- *Applied Bayesian Statistical School.*
- *Scientific board* of BaySM-*Bayesian Young Stitistician Meeting*. *Scientific chair* della prossima edizione.
- *International Society for Bayesian Analysis* - ISBA
- Società Italiana di Statistica-SIS



- *Applied Bayesian Statistical School.*
- *Scientific board* of BaySM-*Bayesian Young Stitistician Meeting*. *Scientific chair* della prossima edizione.
- *International Society for Bayesian Analysis* - ISBA
- Società Italiana di Statistica-SIS

Lucia Paci, Alessia Pini, Raffaele Argiento,

Federico Castelletti, Stefano Peluso, Bruno

Buonaguidi, Guido Consonni

Grazie!