

Premessa

Le premesse dei testi didattico / divulgativi sono sovente caratterizzate dall'elencazione delle motivazioni che hanno indotto gli autori a proporre la propria opera. Cercando allora di soffocare la mia inguaribile natura anarchica che mi rende tendenzialmente allergico ad ogni forma di convenzione, e mettendomi nei panni di uno studente - magari del mio corso - che vorrà utilizzare questo testo, ho sinceramente, e più volte, rovistato nella mia mente alla ricerca di tali motivazioni. Purtroppo ancora oggi non riesco ad individuare un motivo prevalente. Perché in effetti i motivi sono vari e, per dirla in gergo "tecnico", quasi tutti equamente probabili (come in un esperimento casuale ..., ma solo perché ancora non ho deciso quale tra questi è quello che alla fine ha prevalso).

Prima di provare ad elencarli, è bene cercare di definirne l'ambito scientifico di riferimento, data la sua assoluta rilevanza in tutto ciò che caratterizza l'epoca che stiamo vivendo. La velocità con cui si trasferiscono oggi le informazioni sul web richiede la predisposizione di strumenti di analisi sempre più adeguati a trattare moli di dati, di algoritmi sempre più veloci che permettano ai cosiddetti *decision maker* di operare scelte basate su informazioni che il trascorrere del tempo (non più degli anni, ma forse dei giorni se non addirittura delle ore) può rendere obsolete molto velocemente. Con l'avvento di questo secolo si è, non a caso, assistito al susseguirsi della diffusione di nuove terminologie e tecniche di gestione e analisi delle informazioni (spesso strettamente connesse tra loro) quali: *Knowledge Discovery in Databases, Data-warehousing Solutions, Machine Learning, Data Mining, Big Data analytics* ecc.

In questo vasto panorama, appare evidente come gli statistici assumano un ruolo rilevante se non strategico. La trasposizione dei metodi e modelli statistici classici alle opportunità offerte dai moderni calcolatori ha ampliato il ventaglio delle prospettive di analisi e delle possibili applicazioni. Molte di queste metodologie sono superficialmente identificate come appartenenti al generico ambito della Statistica Computazionale, ovvero all'interfaccia

tra la Statistica e l'Informatica. Superficialmente appunto, perché nella letteratura anglosassone si ritiene opportuno compiere uno sforzo semantico che noi trascuriamo, volendo distinguere cosa è inquadrabile come *Computational Statistics* da ciò che invece riguarda lo *Statistical Computing*, ovvero l'applicazione del calcolo elettronico (anche intensivo) ai tradizionali metodi e strumenti propri della Statistica. Nell'ambito della *Computational Statistics* rientra invece tutto ciò che riguarda la progettazione e lo sviluppo di algoritmi finalizzati ad implementare metodi statistici con (e per) i computer, o a risolvere problemi complessi se non impossibili da trattarsi dal punto di vista analitico (Albert e Gentle, 2004; Lauro, 1996; Wilkinson, 2008). Mentre tutti gli statistici dovrebbero mostrare una qualche abilità nell'utilizzo dei più comuni pacchetti statistici, le competenze nell'ambito della *Computational Statistics* non sono per niente scontate.

Molti testi didattici (se in lingua italiana, spesso frutto di un'opera di traduzione) nonostante facciano genericamente riferimento all'ambito della Statistica Computazionale, in realtà si limitano ad introdurre all'utilizzo di un particolare software o pacchetto per l'applicazione dei principali metodi statistici per l'analisi dei dati (per cui dovrebbero essere catalogati come testi di *Statistical Computing*). Mancava a mio avviso un testo che introducesse alla vera arte della Statistica Computazionale, ovvero in grado di illustrare come le competenze di programmazione informatica nello sviluppo di algoritmi possano essere messe al servizio della Statistica per la simulazione e replicazione virtuale di realtà ed esperimenti più o meno complessi. Nella realtà attuale, saper simulare scenari decisionali complessi potrebbe rivelarsi di fondamentale importanza nell'assicurarsi un vantaggio nei confronti dei reali o potenziali competitors.

Ci sono centinaia di testi sul pacchetto **R**. Questo non deve stupire: negli ultimi anni il numero di utilizzatori di **R** al mondo è cresciuto in maniera esponenziale, perché stiamo parlando di un pacchetto estremamente versatile per l'analisi statistico-matematica ed il *Data Mining*, realizzato attorno ad un linguaggio di programmazione vettoriale *java-like*. Il linguaggio quindi ben si presta a svolgere tutte quelle attività di ricerca e analisi che definiscono l'ambito della statistica computazionale. Mi sembra però che siano (in proporzione si intende) relativamente pochi i testi che si occupano di introdurre all'arte della simulazione. E non con il livello di esemplificazione che questo testo intende fornire.

Non verranno pertanto volutamente presentate le funzioni proprie dell'analisi statistica multivariata e per la stima dei parametri dei modelli più comuni. Sarà invece dedicata attenzione all'implementazione di algoritmi per la realizzazione di effetti grafici avanzati, altro elemento a mio avviso carente nei testi didattici che introducono a **R**, che ha appunto nella capacità di sviluppo grafico uno dei suoi punti di forza. Il tutto non dimenticando che quest'opera deve supportare lo studio di un insegnamento da (circa) 9 Crediti Formativi Universitari nell'ambito di un corso di laurea triennale in Statistica.

Sono un perito matematico-informatico come formazione superiore ed uno statistico come formazione universitaria. Ho iniziato a prendere dimestichezza con **R** alla fine del mio percorso di laurea magistrale e solo per condurre analisi finalizzate, e certamente **R** non era quello che è oggi. Quasi tutto quello che verrà qui presentato l'ho quindi auto-appreso in molti anni di studio e lavoro, ed in corsi di aggiornamento certificati dalla John Hopkins University, che mi sono divertito a svolgere grazie all'ottima piattaforma Coursera.

Da studente un testo così mi avrebbe fatto certamente comodo, ma **R** era ancora agli albori. Ecco... forse è questo il motivo principale per cui ho deciso di prendere carta, penna e calamaio (metafora molto poetica... in realtà come immaginerete ho aperto Word).

Buono studio!

fine estate 2017, a Vaglia

