La distribuzione dei cognomi come strumento per l'analisi sociale: l'esempio della docenza universitaria.

Paolo Rossi – Dipartimento di Fisica – Università di Pisa

L'indagine quantitativa sulla distribuzione in frequenza dei cognomi non è soltanto strumento per una miglior comprensione delle dinamiche migratorie o per lo studio dell'evoluzione genetica di una popolazione[1]: è in effetti possibile concepire ed effettuare applicazioni degli esiti di tale indagine all'analisi di alcune dinamiche sociali anche complesse.

A titolo di esempio e anche per l'interesse intrinseco dei risultati ottenuti mostreremo come lo studio della frequenza dei cognomi nella popolazione dei docenti universitari italiani, per la quale esiste una base di dati facilmente accessibile in rete, permetta di ricavare significative indicazioni sull'esistenza e l'eventuale estensione del fenomeno del cosiddetto "nepotismo accademico", almeno nella misura in cui tale fenomeno si manifesti nella forma di un'anomala presenza di relazioni di parentela tra i soggetti investigati.

Notiamo in primo luogo (Tavola I) che la distribuzione dei cognomi all'interno della docenza universitaria (ovvero il numero di cognomi presenti un certo numero di volte nella popolazione di riferimento) quando analizzata empiricamente (alla data del 31 dicembre 2007, dopo la quale non sono avvenute variazioni di rilievo nel reclutamento della prima e della seconda fascia), anche analizzata separatamente secondo le tre fasce (ordinari, associati e ricercatori) mostra l'evidenza di una legge di scala del tipo già più volte osservato, fin dai primi anni Ottanta del secolo scorso[2], in tutte le popolazioni di Stati e Regioni europee presi in esame dai ricercatori.[3] Quest'evidenza è la necessaria premessa per l'utilizzo delle tecniche d'analisi sviluppate in generale per lo studio di fenomeni naturali e sociali caratterizzati da leggi di scala.[4]

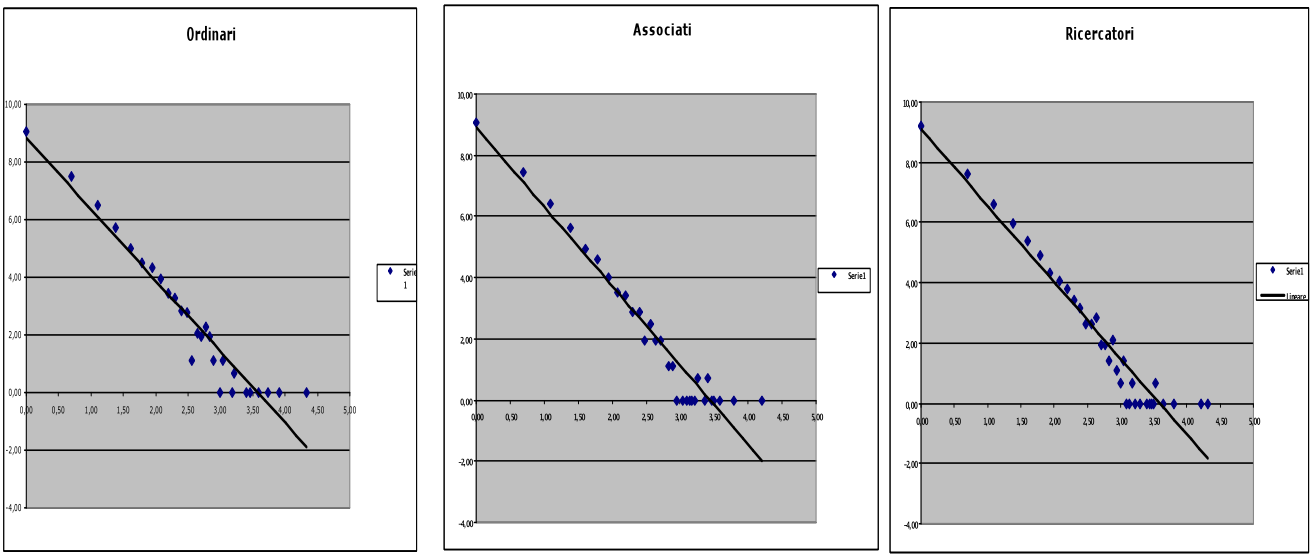

Tavola I – Distribuzione in frequenza dei cognomi nelle tre fasce docenti (scala bi logaritmica)

[1] P. Rossi, *La distribution des noms de famille comme outil pour l'analyse des dinamiques migratoires,* in "Un juego de engaños. Nombres, apellidos y movilidad en los siglos XV al XVII, Madrid 2010
[2] W.R. Fox and G.W. Lasker, *The Distribution of Surname Frequencies*, Int. Stat. Review 51 (1983) 81-87
[3] C. Scapoli, E. Mamolini, A. Carrieri, A. Rodriguez-Larralde, I. Barrai, *Surnames in Western Europe: A comparison of the subcontinental populations through isonymy*, Theor. Pop. Biol. 71 (2007) 37-48
[4] M.E..J. Newman, *Power laws, Pareto distributions and Zipf's law,* Contemporary Physics 46, 323-351 (2005)

Questo specifico studio ha potuto effettuarsi anche grazie all'utilizzo di alcuni concetti che sono stati recentemente introdotti[5] nel contesto dello studio delle distribuzioni in frequenza dei campioni di una popolazione. Risulta *a priori* ovvio che le frequenze misurate sulle distribuzioni di campioni casuali sono connesse, con relazioni determinate dal calcolo delle probabilità, alle frequenze presenti nell'intera popolazione, ma dovrebbe essere parimenti evidente che, quando confrontiamo due campioni differenti, troveremo che in generale, soprattutto per piccoli campioni o per "tipi" scarsamente presenti, un particolare "tipo" non compare con la stessa frequenza nei due campioni.
In particolare, data una popolazione di N elementi e campioni di n elementi ciascuno, possiamo calcolare la correlazione tra due campioni e, se la formazione dei campioni fosse realmente casuale, dovrebbe valere per la correlazione C la relazione

$$C = n/N * (\alpha + N)/(\alpha + n),$$

dove il parametro α (che nel nostro caso misura l'inverso dell'isonimia) è una proprietà dell'intera popolazione, e non dovrebbe modificarsi nel caso di campioni casuali, anche relativamente piccoli.
Come si vede facilmente, la correlazione, come è ovvio che sia, tende a 1 quando n tende a N, ossia quando i campioni coincidono con l'intera popolazione, ma tende a 0 al decrescere di n.

Nel caso dei docenti universitari possiamo studiare separatamente le distribuzioni dei cognomi di ricercatori, associati e ordinari, considerando, come ipotesi di lavoro, che ognuno dei tre gruppi sia semplicemente un campione casuale della popolazione italiana. Lo scostamento dei valori empirici di correlazione da quelli teorici sopra indicati potrà essere interpretato come una misura della non casualità dei campioni (nepotismo accademico). Alcuni risultati sono presentati nella Tavola II, e possono essere interpretati notando che α (inverso dell'isonimia) vale circa 6.000, mentre n per ognuno dei tre campioni vale circa 20.000, per cui il valore atteso di C per campioni non correlati è prossimo a 0,75. Notiamo subito che i valori empirici delle correlazioni, riportati in Tavola II, sono largamente compatibili con l'ipotesi che i tre gruppi di docenti rappresentino campioni casuali, indipendenti tra loro, della popolazione italiana.

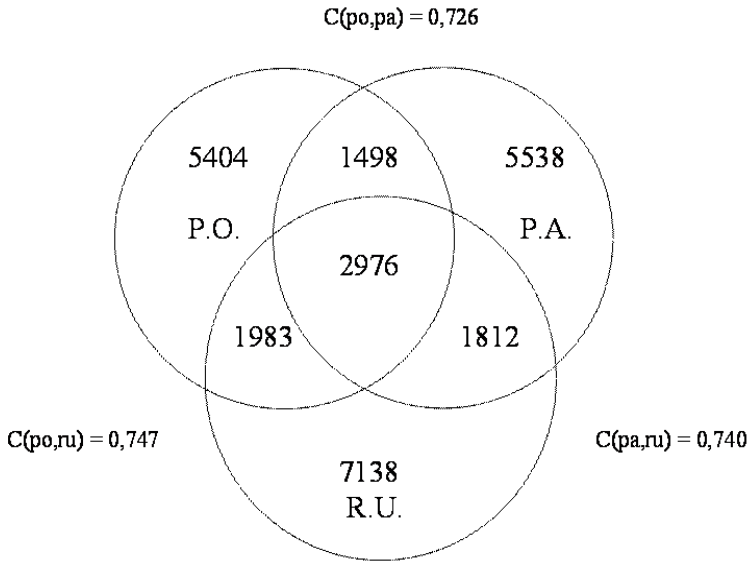

Tavola II – Numero dei cognomi distinti presenti nelle tre fasce, numero dei cognomi presenti in due o più fasce, valore della correlazione tra le distribuzioni.

[5] P. Rossi, *On sampling and parametrization of discrete frequency distributions* (in preparazione)

	ITALIA (elenchi tel.)	UNIVERSITÀ (modello)	UNIVERSITÀ (dati)
INDIVIDUI	**5.031.000**	**62.697**	**62.697**
COGNOMI	**215.623**	**26.408**	**26.562**
FRÉQUENZA 1	**72.450 (33,6%)**	**16.953 (64,2%)**	**16.304 (61,4%)**
FRÉQUENZA 2	**24.000 (11,1%)**	**4.122 (15,6%)**	**4.613 (17,3%)**
FRÉQUENZA 3	**14.450 (6,7%)**	**1.743 (6,6%)**	**1.992 (7,5%)**
ISONIMIA	**0,00017**	**0,00017**	**0,00016**

Tavola III – Valori dei principali parametri teorici ed empirici relativi alla distribuzione.

I dati riportati nella Tavola III sono invece relativi all'intera docenza universitaria. Da un lato essi tendono a rafforzare l'ipotesi sopra enunciata, in quanto il numero totale dei cognomi e l'isonimia riscontrati empiricamente risultano largamente coerenti con i valori di modello, ottenuti postulando l'assoluta casualità del campione.
D'altra parte invece l'analisi dettagliata delle frequenze mostra una leggera, ma non completamente trascurabile, deviazione dalle previsioni statistiche. In particolare il numero dei cognomi presenti una sola volta nella distribuzione è lievemente inferiore alla previsione teorica, mentre il numero dei cognomi presenti più di una volta è un poco superiore alla previsione. Sono almeno cinquecento più del previsto i cognomi presenti due volte anziché una sola, e tra i cognomi che compaiono tre volte ve ne sono almeno centocinquanta in eccesso.
Si potrebbe pertanto concludere che il "nepotismo accademico", almeno per la parte misurata dalle relazioni di parentela diretta in linea paterna, e quindi associabile alla ripetizione del cognome, è un fenomeno non drammatico sul piano statistico generale, in quanto l'ipotesi che la docenza nel suo complesso rappresenti un campione casuale dell'intera popolazione italiana è sostanzialmente verificata, ma ciò nonostante non si tratta di un fenomeno del tutto trascurabile, in quanto la ripetizione dei cognomi è sicuramente più elevata, ancorché di pochi punti percentuali, rispetto a quanto ci si sarebbe aspettato a livello teorico e modellistico.

A suffragio di quest'interpretazione viene anche la ricerca recentemente svolta dal dott. G. Daniele, che nella sua tesi di laurea[6], proprio partendo da una dettagliata analisi della distribuzione dei cognomi, effettuata a livello di singoli Atenei e Facoltà, ha messo in evidenza che il fenomeno del nepotismo, nel senso qui indicato, è soprattutto peculiare di alcune ben individuate Sedi e Facoltà universitarie, situate in larga prevalenza nel Meridione d'Italia. Questa constatazione sembra fornire una sufficiente spiegazione della marginalità statistica complessiva del fenomeno ma anche della presenza di un numero di ripetizioni di cognomi significativamente superiore ai valori attesi,
Sul piano della sociologia della ricerca è interessante ricordare che dal lavoro del dott. Daniele emerge anche un'importante correlazione negativa tra il nepotismo accademico, là dove osservato e misurato, e la qualità della ricerca, misurata mediante gli indicatori bibliometrici di uso più comune.

La sempre maggior disponibilità di informazioni sulla composizione presente e passata della docenza universitaria italiana e sulla sua produzione scientifica potranno costituire il punto di partenza per più approfondite indagini, anche di tipo diacronico, sui fenomeni e sui processi che abbiamo in questa sede preso sommariamente in esame.

[6] G. Daniele, *L'Università pubblica italiana: qualità e omonimia tra i docenti*, Tesi (rel. Prof. N.D. Coniglio), Facoltà di Economia, Università degli Studi di Bari (2010)