Un progetto di analisi statistica dei dati genealogici relativi a Montecarlo di Lucca in età moderna

Sergio Nelli*, Paolo Rossi**, Roberto Bizzocchi**

* *Archivio di Stato di Lucca;* ** *Università di Pisa*

Il lavoro di ricerca già effettuato, incentrato sulla raccolta, schedatura e analisi dei dati anagrafici contenuti nei registri parrocchiali e nel materiale d'archivio disponibile per il territorio di Montecarlo (LU)[1], facendo riferimento a un arco di tempo che supera i confini dell'intera età moderna, costituisce il fondamentale punto di partenza per un ulteriore progetto di ricerca che abbiamo recentemente avviato e che qui si vuole sommariamente descrivere.

La prima fase del progetto consiste nella creazione di una base di dati strutturata e facilmente accessibile via *web*, che a partire dalla documentazione raccolta la organizzi in forma di schede individuali, rintracciabili secondo differenti chiavi di ricerca (nome, cognome, date di nascita, matrimonio e morte, eventuali ulteriori elementi identificativi), e collegate tra loro mediante *links* esplorabili in modo automatico.

Confrontandoci con modelli già presenti in rete (usualmente utilizzati per la rappresentazione delle relazioni familiari della nobiltà europea)[2], e grazie alle competenze già localmente acquisite in questo campo, sappiamo che una base di dati come quella sopra descritta potrà permettere in particolare la ricostruzione e la generazione automatica, anche in forma grafica, di tutti gli alberi genealogici (e delle *Ahnentafeln*) già implicitamente contenuti nella documentazione.

Questa parte del progetto, anche a prescindere dal suo intrinseco interesse a livello documentario, sarà comunque funzionale a due ulteriori linee di ricerca che è nostra intenzione portare avanti, dapprima

[1] S. Nelli, **??**.

[2] Cfr. ad esempio i siti: http://roglo.eu/roglo e http://www8.informatik.uni-erlangen.de/html/ww-person.html

in modo indipendente, per poi tentare un confronto tra i risultati ottenuti che potrebbe risultare illuminante, come spiegheremo più avanti, al fine di verificare con un esempio concreto e sufficientemente articolato alcune ipotesi presenti, in particolare nella letteratura relativa alla genetica delle popolazioni e alla demografia storica.

La prima linea di ricerca è propriamente collegata alle problematiche che nascono dallo studio dell'origine storica e della diffusione dei cognomi. Si tratta qui di indagare, in modo anche quantitativo, i differenti processi che portano all'evoluzione del patrimonio di cognomi presenti in un territorio, e che possono in estrema sintesi essere ricondotti a tre tipologie fondamentali[3]:

– la *differenziazione* per cui differenti rami originati da uno stesso ceppo familiare possono assumere dopo qualche generazione denominazioni diverse onde favorire un più preciso riconoscimento di gruppi di individui ormai separati sia negli interessi materiali che nella collocazione fisica; un sottocaso della differenziazione è la *mutazione* della pronuncia e della forma grafica del cognome;
– l'*immigrazione*, per cui nuovi cognomi compaiono un territorio provenendo da aree differenti a causa dello spostamento permanente di individui e gruppi familiari;
– l'*estinzione*, dovuta all'assenza o alla sterilità di tutti i discendenti maschi, un fenomeno che ha una calcolabile (e non trascurabile) probabilità, e che in comunità a crescita limitata porta, su scala secolare, alla scomparsa di una frazione importante dei cognomi inizialmente presenti.

Esistono modelli matematici, ormai largamente studiati, volti a mettere in relazione i parametri derivabili empiricamente dall'analisi sopra citata (frequenza delle mutazioni, coefficiente d'immigrazione, probabilità di estinzione) con la distribuzione in frequenza dei cognomi nella popolazione (ovvero con il numero di cognomi rappresentato nella popolazione da un dato numero di individui)[4]. Sarà quindi possibile

[3] 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*, a cura di G. Salinero, I. Testón Núñz, Madrid, Casa de Velázquez, 2010, pp. 153-161.

[4] S.K. Baek, H.A.T. Kiet, B.J. Kim, *Family name distributions: Master equation approach*, in "Physical Review E", 76 (2007) 046113; A. De Luca, P. Rossi, *Renormalization group evaluation of exponents in family name distributions*, in "Physica A" 388 (2009), 17, pp. 3609-3614.

effettuare una verifica della concreta applicabilità di questi modelli, avendo a disposizione per una stessa popolazione sia il valore empirico dei parametri suddetti che l'effettiva distribuzione dei cognomi all'ultima data utile per la quale si sia in possesso dei dati anagrafici rilevanti.

Dovrebbe risultare abbastanza evidente la rilevanza per gli studi di demografia storica di una verifica empirica di questi modelli su una popolazione reale, in quanto ben raramente sarà possibile in altri casi ricostruire su base strettamente documentaria la dinamica effettiva della popolazione, specialmente se essa è vasta, mentre quasi sempre è possibile acquisire una "fotografia" della popolazione a una data certa. Lo studio della distribuzione dei cognomi a una data anche recente, quando si fosse in possesso di un modello consolidato della dinamica che può aver portato a quella distribuzione, permetterebbe una ricostruzione induttiva dei parametri relativi ai processi di differenziazione e di migrazione anche in assenza della relativa documentazione.

La seconda linea di ricerca prende le mosse dal contributo seminale di Derrida *et al.*[5], che hanno formulato nel 1999 una teoria statistica degli alberi genealogici, capace di render conto dell'effettiva molteplicità del numero degli antenati di un singolo individuo sulla base dei fenomeni di *inbreeding* (matrimonio tra parenti, anche lontani) che risultano inevitabili in una comunità chiusa, ma si verificano comunque prima o poi anche nelle comunità più vaste ed aperte. Risulta infatti intuitivo che il numero degli antenati distinti presenti nell'*Ahnentafel* di ciascun individuo non può raddoppiare indefinitamente a ogni generazione, come avviene di regola nelle prime poche generazioni, in quanto tale numero finirebbe ben presto per superare le dimensioni dell'intera popolazione alla data del passato corrispondente a quella generazione.

La base di dati che si renderà disponibile sarà particolarmente adatta a una valutazione dell'effettiva rapidità con cui si sono manifestati nel caso in esame i processi di *inbreeding*, e permetterà una misura quantitativa della consanguineità, ovvero della "parentela media", esistente nella popolazione, ed eventualmente della variabilità nel tempo di tale parametro. L'interesse di tale misura è largamente interdisciplinare, data

[5] B. Derrida, S.C. Manrubia, D.H. Zanette, *Statistical Properties of Genealogical Trees*, in "Physical Review Letters", 82 (1999), 9, pp. 1987-1990; S.C. Manrubia, D.H. Zanette, *At the Boundary between Biological and Cultural Evolution: the Origin of Surname Distributions*, in "Journal of Theoretical Biology", 216 (2002), 4, pp. 461-477; S.C. Manrubia, B. Derrida, D.H. Zanette, *Genealogy in the Era of Genomics*, in "American Scientist" 91 (2003), pp. 158-165.

la sua rilevanza non soltanto per la demografia storica e per la sociologia delle relazioni familiari della popolazione in esame, ma anche per le implicazioni dell'*inbreeding* sulla struttura genetica della popolazione stessa[6]. Queste tematiche sono già state largamente studiate a livello teorico, ma ancora una volta il numero delle verifiche empiriche dei modelli è estremamente limitato, per l'ovvia difficoltà di disporre di basi di dati sufficientemente ampie e documentate nei loro aspetti diacronici.

A conclusione di questo duplice percorso di ricerca è immaginabile un ulteriore sviluppo, la cui effettiva praticabilità dipenderà ovviamente dalla qualità dei risultati ottenuti nella ricerca stessa. Si tratta infatti di effettuare una comparazione dei dati relativi alla misura diretta della consanguineità della popolazione con quelli che possono essere desunti tramite lo studio dei cognomi negli atti matrimoniali, e in particolare con il parametro definito "isonimia matrimoniale", che consiste nella stima della frequenza dei matrimoni nei quali entrambi i coniugi hanno lo stesso cognome[7]. Questo studio, avviato sistematicamente in altro contesto da Cavalli-Sforza[8] e collaboratori fin dagli anni Cinquanta del secolo scorso, è stato messo in relazione diretta, a livello teorico, con l'indice di consanguineità. Poiché nel caso in esame possiamo sperare di giungere a una misurazione concettualmente indipendente di isonimia e consanguineità, sarà possibile, se i dati empirici risulteranno sufficientemente accurati e statisticamente significativi, giungere a un confronto diretto delle due stime e a una validazione (o falsificazione) empirica dell'assunto teorico, con importanti conseguenze su ogni ulteriore utilizzo della metodologia basata sull'isonimia nel contesto della genetica delle popolazioni e della demografia storica.

[6] G.W. Lasker, *Surnames and Genetic Structure*, Cambridge University Press, 1985; M.A. Jobling, *In the name of the father: surnames and genetics*, in "Trends in Genetics", 17 (2001), 6, pp. 353-357; J.P. Darlu, *Patronymes et démographie historique*, in "Annales de Démographie Historique", 108 (2004), 2, pp. 53-65.

[7] F. Crow, A.P. Mange, *Measurement of inbreeding from the frequency of marriages between persons of the same surname*, in "Eugenics Quarterly", 12 (1965), pp. 199-203.

[8] L.L. Cavalli-Sforza, A. Moroni, G. Zei, *Consanguinity, Inbreeding and Genetic Drift in Italy*, Princeton University Press, 2004.