

GARY KING, *A Solution to the Ecological Inference Problem. Reconstructing Individual Behavior from Aggregate Data*, Princeton, Princeton University Press, 1997, pp. 342, £ 45, Isbn 0-691-01241-5 (hb.)

Gary King è docente di Government presso il Department of Government della Harvard University (lo stesso di Verba). Ma è anche direttore dell'Harvard-MIT Data Center e presidente della Society for Political Methodology. Questa accoppiata di sensibilità politica e di specializzazione metodologico-statistica non costituisce certo una novità per le università americane, ma i lavori di King si fanno apprezzare per il completo controllo di entrambe le discipline, quasi a indicare un passaggio generazionale e un salto curriculare di indubbia rilevanza. Il libro qui recensito ne costituisce l'esempio più tipico (anche nei suoi difetti come, ad esempio, l'esclusivo riferimento alla letteratura e al dibattito statunitense); non a caso ha avuto numerosi riconoscimenti, tra cui il premio Gosnell per il migliore lavoro metodologico dell'anno scorso in scienza politica.

L'oggetto del libro di King è ben noto agli «elettoralisti» italiani per le polemiche, durate un ventennio, sulla robustezza delle stime dei flussi elettorali ottenute attraverso procedure di stima che riprendono il modello di Goodman (pur con adattamenti e accorgimenti vari). Ridotto ai suoi termini essenziali, il problema è di sapere «chi ha votato cosa», avendo a disposizione solo dati aggregati, ovvero noti i marginali di riga e di colonna di una matrice-dati ma non i valori delle singole celle. Da noi la disputa ha assunto toni a volte da *querelle* ideologico-filosofica; dall'altra parte dell'oceano la questione ha avuto invece risvolti più pragmatici e venali, dal momento che le stime del «voto razziale» sono accettate come prova negli infiniti ricorsi davanti alle corti sul disegno dei collegi elettorali che violano il Voting Right Act.

Di qui il tono meno enfatico della discussione, documentato anche da questo importante lavoro, alla cui base sta la consapevolezza dell'irrisolubilità del problema della certezza dei risultati. Trattandosi di stime, la questione non è se siano esatte o sbagliate, quanto invece di avere metodi di calcolo che riducano i margini di errore, che stimino l'errore stesso, che producano statistiche in grado di misurare i costi e i benefici degli assunti di volta in volta introdotti negli algoritmi di stima. Fin dal titolo del libro questo approccio realistico è ampiamente enfatizzato: non la, ma una soluzione al problema dell'inferenza ecologica; per poi rimarcare in sede introduttiva come la strada proposta sia passibile di molti altri miglioramenti. King è drastico al riguardo: «siccome il problema dell'inferenza ecologica è causato da un'assenza di informazioni a livello individuale, nessun metodo di inferenza ecologica, compreso quello presentato in questo libro, è in grado di produrre risultati attendibili in ogni circostanza».

Il volume è diviso in cinque parti. Nei due capitoli introduttivi della prima parte si presentano gli antecedenti storici del problema, viene avanzata in forma intuitiva il tipo di soluzione proposta, si introducono le notazioni matematiche utilizzate nel resto del libro. La seconda parte inizia con una importante dimostrazione della sostanziale equivalenza da un punto di vista matematico di tutte le soluzioni proposte al problema del *bias* da aggregazione. Si passa poi a presentare una serie di problemi statistici che si incontrano nell'applicazione del modello di Goodman, che non hanno a che fare con il problema aggregativo. Essi dovrebbero sempre essere valutati quando si vogliono applicare modelli di stima ecologica che finora sono stati ampiamente trascurati (anche se alcuni di essi erano stati sollevati nel dibattito italiano, del tutto sconosciuto a King).

Il cuore del libro è nella parte terza. Viene riformulato il problema aggregativo e proposta una soluzione sia algebrica che grafica per controllare le possibili soluzioni alternative compatibili con i dati a disposizione, presentando poi un modello statistico e probabilistico che consente di ridurre l'incertezza tra le soluzioni possibili. Nei capitoli successivi si sviluppano le procedure di stima del modello e si introducono test di controllo delle operazioni che passo passo vengono compiute. Infine, King propone una dimostrazione di come la sua soluzione sia matematicamente equivalente a quella proposta per risolvere problemi curiosamente simili in geologia o in tomografia: come quando, ad esempio, si devono ricostruire immagini stratificate di parti del corpo umano a partire dalle informazioni dei raggi X.

La quarta parte del volume è dedicata alle verifiche empiriche su basi-dati molto diverse tra loro (registrazione elettorale secondo razza; status secondo sesso; voto secondo razza). L'ultima parte è dedicata alle conclusioni e all'ampliamento delle prospettive di analisi in varie direzioni, compresa quella già sperimentata nel nostro paese di combinare dati di sondaggio e dati aggregati per migliorare l'inferenza ecologica. Da ultimo, si dimostra (solo matematicamente) l'estendibilità dei procedimenti proposti dalle matrici « 2×2 » alle matrici « $n \times n$ ». Il libro è chiaro, ben costruito, convincente. Il programma che implementa tutte le procedure descritte nel volume è libero e facilmente acquisibile. Con un unico difetto: è fatto per gli Stati Uniti, un paese fortunato, dove sembra che tutti i problemi ecologici siano ridicibili a matrici « 2×2 ». Chi invece deve lavorare con matrici di dati con più di quattro celle deve attendere che qualcuno si metta d'impegno a riconfigurare tutti i parametri (e non solo) del programma.

[Paolo Feltrin]