

$$\frac{A_{13}}{378}$$

Vincenzo Marinello

Analisi dei flussi elettorali

Un approccio alla stima mediante un modello
gerarchico bayesiano



Copyright © MMX
ARACNE editrice S.r.l.

www.aracneeditrice.it
info@aracneeditrice.it

via Raffaele Garofalo, 133/A-B
00173 Roma
(06) 93781065

ISBN 978-88-548-3232-9

*I diritti di traduzione, di memorizzazione elettronica,
di riproduzione e di adattamento anche parziale,
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

*Non sono assolutamente consentite le fotocopie
senza il permesso scritto dell'Editore.*

I edizione: maggio 2010

Indice

- 7 Introduzione
- 11 Capitolo I
Cenni sugli approcci socio-politologici
- 17 Capitolo II
Metodi di stima dei flussi elettorali
- 2.1. Estrazione dell'informazione deterministica: il metodo dei *bounds*, 20 – 2.2. Estrazione statistica dell'informazione: il modello di regressione di Goodman, 23 – 2.3. Estrazione deterministica e statistica dell'informazione: il metodo di King, 26
- 33 Capitolo III
Il modello gerarchico bayesiano di King, Rosen e Tanner
- 3.1. Il modello gerarchico bayesiano, 33 – 3.1.1. Il modello gerarchico bayesiano per tabelle 2×2 , 35 – 3.1.2. Il modello gerarchico bayesiano di King, Rosen e Tanner per tabelle $R \times C$, 38 – 3.2. I metodi computazionali MCMC, 42 – 3.2.1. Il campionamento di Gibbs, 45 – 3.2.2. Il metodo di Metropolis-Hastings, 47 – 3.2.3. Algoritmi composti, 52 – 3.2.4. Diagnosi della convergenza e analisi dell'*output* per MCMC, 53

57 Capitolo IV
Un'applicazione: i flussi elettorali 2006-2008 per la
Camera dei Deputati nel Comune di Sciacca

4.1. Richiami e considerazioni sul sistema elettorale, 57 – 4.2. L'unità di analisi e le variabilità d'interesse, 58 – 4.2.1. La categorizzazione della variabile "Lista elettorale", 61 – 4.3. Analisi descrittive sui dati elettorali del Comune di Sciacca, 67 – 4.3.1. Analisi descrittive sulle distribuzioni marginali, 69 – 4.4. Applicazione del modello gerarchico bayesiano per la stima dei flussi elettorali relativi alle elezioni del 2006 e 2008 per la Camera dei Deputati del Comune di Sciacca, 73

85 Conclusioni

87 Elenco delle tabelle

89 Elenco delle figure

91 Bibliografia

Introduzione

L'argomento di questo lavoro è la stima dei flussi elettorali a partire dai risultati effettivamente registrati in due elezioni successive dello stesso tipo. Si tratta di un problema statistico che si inquadra nel più ampio tema dello studio dell'evoluzione del comportamento del corpo elettorale, un tema di grande interesse politico e sociale, come testimonia il fatto che il mondo della comunicazione ha trasformato il dibattito sulle opinioni politiche dei cittadini in un vero e proprio "spettacolo", che alimenta per giorni le cronache dei mass media, soprattutto nel periodo successivo alle votazioni. Il tema sembra essere meno sentito dalle istituzioni pubbliche; tuttavia, sarebbe invece di grande significato un'attività di indagine svolta da un ufficio pubblico di statistica. Infatti, fornire informazione alle amministrazioni ed ai cittadini sui comportamenti elettorali è in effetti un modo per rafforzare il consenso nei confronti delle istituzioni, consenso che si intende con riferimento alla capacità di rappresentanza degli interessi sociali, quindi alla tenuta democratica delle istituzioni medesime.

Chiarita la cornice nella quale il lavoro si colloca, è utile introdurre i metodi attualmente utilizzati per ottenere informazioni sulle preferenze politiche degli elettori e sulla loro evoluzione. Per studiare le variazioni dei comportamenti elettorali nei confronti delle coalizioni e/o dei partiti politici esistono, in letteratura, diversi metodi. Tali metodi si possono classificare in base all'approccio che li contraddistingue.

Un primo approccio, basato sul contatto diretto con l'elettorato, produce i suoi risultati ricorrendo a interviste: *sondaggi* ed *exit-polls*. Questi metodi permettono di ottenere delle informazioni sia prima che in seguito all'elezione e di inferire sul comportamento dell'elettorato utilizzando dati individuali. È tuttavia possibile che i dati risultanti dalle interviste non corrispondano al dato ufficiale. Questo rischio è particolarmente evidente nelle interviste condotte mesi prima delle elezioni quando l'elettore o è incerto o ha ancora la possibilità di cambiare opinione. Recenti insuccessi dei risultati di indagine condotte con *exit polls* dimostrano che questo approccio soffre di grosse limitazioni, specialmente in un paese, come l'Italia, che tende ad essere reticente nel rivelare i propri orientamenti politici.

Un secondo tipo di approccio affronta il tema in termini macro, guardando al comportamento di aggregati omogenei del corpo elettorale individuati, a seconda dell'approccio, nell'appartenenza degli elettori a determinati gruppi politici o sociali. Questo approccio ha il grande vantaggio di potere contare sulla disponibilità (anche per i livelli territoriali subcomunali) di molteplici basi di dati sociali e demografiche e sul consistente patrimonio di informazioni politiche contenute nelle basi di dati dei risultati elettorali che sono disponibili a livello di singola sezione elettorale. In termini metodologici il problema è quello di collegare le informazioni elettorali disponibili a livello di una singola sezione per una certa votazione (i risultati elettorali sezionali) con le informazioni relative alla composizione socio-demografica dell'elettorato: in questo caso, si parla di *analisi della composizione di classe del voto*. Si possono inoltre collegare i risultati elettorali delle elezioni precedenti con quelle delle elezioni successive in tal caso si parla di *analisi dei flussi elettorali*.

Il problema dello studio dei flussi elettorali può essere formalizzato trattando i dati disponibili come distribuzioni marginali di un'ipotetica tabella a doppia entrata che rappresenta il comportamento di voto. Le celle della tabella sono interpretabili come flussi direzionali, vale a dire elettori che hanno votato la "Lista P" in una tornata elettorale e la "Lista Q" nella tornata successiva. Inizialmente le celle della tabella sono vuote, e il problema metodologico e tecnico consiste proprio nel trovare adeguate strategie di riempimento delle celle introducendo opportune ipotesi. Il contenuto delle celle viene usualmente stimato attraverso metodi statistici che hanno in comune il fatto di trattare i risultati elettorali per singola sezione come "realizzazioni casuali" di un qualche processo ignoto, ma reale. Il vantaggio di questo approccio è quello di avere una quasi perfetta corrispondenza dei risultati delle analisi con i dati ufficiali. Lo svantaggio è quello di rendere più complessa l'inferenza sul comportamento individuale dell'elettore. L'unità di osservazione non è più l'elettore ma l'unità aggregata rappresentata, ad esempio, dalla sezione o dalla circoscrizione elettorale. Si cercherà, dunque, di studiare il comportamento individuale a partire dai dati aggregati, un tipo di analisi statistica che va incontro al noto problema della "fallacia ecologica", cui si farà riferimento nel proseguo del lavoro. È bene ribadire che il processo a cui si fa riferimento è di natura prettamente sociale, anche se viene modellato in termini matematici, e che, quindi, i flussi stimati hanno un significato non solo statistico, ma politico dal momento che rappresenterebbero la stima di una sorta di comportamento medio del corpo elettorale.

La struttura del lavoro è la seguente. Nel primo capitolo, si daranno dei cenni sugli aspetti socio-politologici del fenomeno. Nel secondo capitolo, si parlerà dei metodi di stima dei flussi elettorali presenti in letteratura.

Nel terzo capitolo, si discuterà del modello gerarchico bayesiano Multinomiale-Dirichlet e si illustreranno le tecniche computazionali. Nel quarto capitolo si svolgerà un'applicazione sui dati elettorali del Comune di Sciacca. In particolare, si analizzeranno i risultati dell'analisi condotta con il modello gerarchico bayesiano Multinomiale-Dirichlet. Chiudono il lavoro le conclusioni, dove si riassumeranno i risultati del lavoro svolto.

Cenni sugli approcci socio-politologici

Le analisi del comportamento elettorale hanno a lungo sottolineato il radicamento dei partiti in alcuni gruppi sociali, e, al contempo, la loro capacità di creare rapporti forti di identificazione con l'elettorato. Queste tesi sono state in primo luogo sostenute dalle due principali scuole sviluppate negli Stati Uniti, attorno agli anni Quaranta, cercando di spiegare il comportamento elettorale: l'una detta *Scuola di Colombia*, ha sottolineato il ruolo economico di appartenenza; l'altra, la *Scuola di Michigan*, ha guardato piuttosto alle caratteristiche psicologiche individuali. In entrambi i casi, si è osservata la durezza dell'attaccamento ai partiti attraverso il voto che si può definire di "appartenenza".

Secondo la Scuola di Colombia, nata attorno ai progetti di ricerca guidati da Paul Lazarsfeld, il voto, pur essendo un comportamento individuale, è però profondamente influenzato dalle norme e dai valori dominanti nei diversi gruppi sociali. Molto spesso la scelta di voto può essere fatta risalire alla socializzazione politica in famiglia, dato che i figli votano (nel 77% dei casi studiati da Lazarsfeld) per lo stesso partito per il quale votano i genitori. La conclusione di Lazarsfeld è che "una persona pensa politicamente come è socialmente". Le caratteristiche sociali determinano le scelte politiche. Status socio-economico, religione e luogo di residenza sono i principali predittori della decisione di voto, che tende poi a mantenersi nel tempo.

I ricercatori della cosiddetta Scuola di Michigan, analizzando una serie storica di sondaggi su campioni molto vasti, hanno sottolineato alcune caratteristiche di psicologia individuale che prevarrebbero rispetto all'appartenenza sociale nel determinare la scelta dell'elettore. Il comportamento di voto sarebbe secondo questa scuola influenzato da una serie di atteggiamenti individuali su tematiche, candidati e programmi, e dalla intensità di questi atteggiamenti. Anche secondo questa scuola, comunque, l'adesione ad una serie di valori e posizioni interagisce con l'identificazione con un partito, che tende ad essere precoce e coerente con quella dei genitori. Questa identificazione tende a mantenersi stabile (per due elettori su tre è costante nelle diverse elezioni), e a rafforzarsi con il passare del tempo (più l'elettore invecchia, più cresce l'intensità dell'identificazione con il partito, riducendosi quindi la probabilità di cambiare).

Successivamente, si sono comunque osservate una riduzione dell'elettorato di appartenenza – legato ad una identificazione di lungo periodo con i partiti – ed una crescita invece degli elettori cosiddetti indipendenti, disponibili a farsi convincere a cambiare partito.

La ricerca contemporanea indica che, mentre l'elettorato di appartenenza – lo “zoccolo duro” dei partiti – ha teso a sfaldarsi, è aumentato l'elettorato di opinione, che vota sulla base di specifiche preferenze per i programmi dei partiti sui singoli temi, mutando anche partito da un'elezione all'altra, mentre persiste un elettorato di scambio, cioè di elettori che scambiano il loro voto con favori.

Diverse ricerche hanno evidenziato, inoltre, un drastico declino del voto di classe, sul quale avevano insistito Lazarsfeld e la Scuola di Columbia. Se il peso dell'appartenenza sociale rispetto al voto non è stato del tutto superato (ad esempio, i partiti di centro-destra continua-

no ad avere i loro principali bastioni fra gli elettori dei ceti medi e alti fra chi vive in campagna o in comuni di piccole dimensioni, e fra gli elettori più religiosi), molte ricerche indicano comunque che l'influenza della classe sociale di appartenenza sul comportamento elettorale è, nel mondo occidentale, in declino. Si deve aggiungere che non sono più solo i meno interessati alla politica a cambiare. Come ha rilevato Dalton, si è ridotto l'attaccamento ai partiti sia fra coloro che non hanno interesse alla politica, che fra chi mantiene quell'interesse. Si potrebbe quindi pensare che, insieme all'istruzione e alla modernizzazione culturale, siano aumentati gli elettori critici che giudicano prima di votare. L'elettore è stato infatti presentato dai più recenti approcci razionali alla politica come capace di valutare la *performance* dei politici e di agire di conseguenza. L'elettore razionale sa giudicare i partiti e scegliere quelli che massimizzano i suoi desideri. In generale, egli valuterà sulla base delle sue preferenze il rendimento del partito al governo, votandolo se trova che la situazione è migliorata nel periodo in cui è stato al governo, o votando per l'opposizione nel caso inverso. L'elettore non identificato è, per definizione, un elettore più volatile. Se Lipset e Rokkan avevano sottolineato la continuità nei sistemi di partito, molti studi recenti hanno messo in discussione, o almeno delimitato, la validità dell'ipotesi di congelamento delle fratture. Già a proposito degli anni ottanta, si è cominciato a parlare di una liberazione di molti sistemi di partito dalla "camicia di forza delle tradizionali fratture". Scongelamento vuol dire innanzitutto indebolimento della fedeltà ai partiti tradizionali e aumento della volatilità elettorale. Con l'effetto dello scongelamento, sembra inoltre cresciuto l'astensionismo elettorale.

Anche nel panorama Italiano sono presenti diversi approcci allo studio dei comportamenti elettorali. Come

in tutti i “regimi democratici” i sondaggi elettorali svolgono un ruolo centrale anche nel nostro Paese, in quanto permettono di conoscere, o almeno così si crede, gli orientamenti dei cittadini nei confronti di istituzioni, partiti, *leader* politici e politiche pubbliche. I risultati dei sondaggi godono di una visibilità mediatica sempre più marcata, il che però rende più effimero il loro impatto. Gli attori politici se ne servono per impostare le loro strategie di comunicazione e per tentare di influire sull’opinione pubblica (o persino di manipolarla). Un numero elevato di aziende danno vita a un mercato concorrenziale per la conduzione dei sondaggi, con il rischio di impoverire la qualità delle ricerche pur di abbassarne i costi. Le diverse categorie di utenti di queste inchieste, infine, difficilmente posseggono le competenze metodologiche per valutarle criticamente. Vista la rilevanza delle funzioni dei sondaggi e dei rischi connessi, è essenziale che i loro risultati rispecchino fedelmente gli orientamenti dei cittadini e che vi trovino riscontro le relative interpretazioni. In Italia, esistono molti Istituti di sondaggi tra i quali ad esempio l’Istituto Cattaneo, Ekma Ricerche, IPR Marketing, FRENI RICERCHE, IPSOS, Eurispes, DIGIS, E-RES, ISTAT, Korus, Eurema, IPR, Eurobarometer, Coesis Research.

Altra forte caratterizzazione del sistema Italia è sicuramente data dall’attività “politica” dei media. Una lunga tradizione di ricerca ha affrontato il problema degli effetti dei media, specialmente la televisione, sulle scelte di voto, giungendo a conclusioni spesso controverse. La comunicazione interpersonale, a cui i primi studi statunitensi sulle campagne elettorali attribuiscono un ruolo cruciale, ha spesso ricevuto minore attenzione, specie nella ricerca italiana. Alcuni studi a tal proposito possono essere trovati nel sito della Società Italiana di Studi Elettorali (SISE: www.studielettorali.it).

Altra componente di notevole importanza è la relazione tra i risultati del voto elettorale e la condizione socioeconomica. Diversi studi hanno evidenziato anche in Italia come l'appartenenza sociale costituisca una variabile fondamentale per la comprensione degli orientamenti di voto fino a tutti gli anni '80, per poi diminuire drasticamente in occasione delle politiche del 1992 fino ai nostri giorni. Alcuni articoli su questi temi possono essere reperiti nel sito della SISE.

Tutti questi aspetti socio-politologici nello studio del comportamento elettorale, nel contesto Italiano, trovano spazio su riviste come: *il Mulino*, *Il Politico*, *Polis*, *Quaderni di Scienza Politica*, *Rassegna Italiana di Sociologia*, *Rivista Italiana di Scienza Politica*, *Teoria Politica*.

Metodi di stima dei flussi elettorali

I primi studi volti alla stima dei flussi elettorali, per citare i più famosi, furono affrontati da Bernstein negli anni '30, da Duncan, Davis e Goodman negli anni '50, proseguirono con Blalock negli anni '60 con il *modello dell'urna* per finire con i più recenti dovuti a King, Rosen e Tanner. Bernstein introdusse il *modello di regressione ecologica* ed anticipò Robinson sul problema della *fallacia della correlazione ecologica*. I primi modelli erano pensati per l'analisi delle transizioni in sistemi bipartitici anche se Goodman, in seguito, sviluppò un modello per i sistemi multipartitici. Per molti anni il modello di Goodman ha riscosso grande fortuna perché grazie alla sua semplicità permette di effettuare studi esplorativi accettabili in termini di tendenze. Dal 1953 fino al 1997, i due approcci largamente usati nel problema della stima dei flussi elettorali furono il *metodo dei bounds* e la *regressione di Goodman*. King combinò l'informazione dei *bounds* con l'approccio di Goodman per estrarre informazione all'interno dei *bounds*.

King, Rosen e Tanner svilupparono un modello gerarchico a tre stadi basato su un approccio bayesiano per cercare di utilizzare l'informazione a priori sui comportamenti individuali combinando tali informazioni con i dati aggregati. L'approccio consente, inoltre, di inserire all'interno del modello delle variabili esplicative che possano incrementare l'informazione sulla variabile di risposta.

Altro approccio è quello di Thomsen dell'Università di Aarhus, nel 1987 propose un metodo di stima dei flussi elettorali basato su *modelli logit* e di *latent structure analysis*, con l'intenzione di superare il problema della fallacia ecologica.

Nel tentativo di risolvere questo problema, ci si trova a doverne affrontare uno più complesso: quello ecologico, noto come "problema di Robinson" o della "*ecological fallacy*". Con questo termine ci si riferisce all'arbitrarietà della verifica delle relazioni tra variabili a livello individuale mediante l'uso di correlazioni su variabili rilevate a livello aggregato. Una correlazione individuale e una correlazione in cui l'unità è indivisibile e le variabili osservate sono proprietà descritte dagli individui. Una correlazione ecologica è una correlazione in cui l'unità è un gruppo di persone accomunate da una determinata caratteristica, come l'appartenenza ad una determinata area geografica, e le variabili osservate sono percentuali o medie, proprietà descrittive dei gruppi. In casi estremi le relazioni aggregate potrebbero essere inverse rispetto a quelle individuali, un fenomeno molto simile al paradosso di Simpson (Simpson, 1951). Ad esempio, si consideri la relazione tra nativi e non analfabeti discussa da Robinson. Per ognuno dei 48 stati USA (1930), Robinson calcolò due proporzioni: la percentuale della popolazione di nati all'estero e la percentuale di popolazione non analfabeta. La base di ogni proporzione è il numero di residenti dello stato che hanno compiuto almeno 10 anni. La correlazione tra le 48 coppie di numeri è 0.53. Questa è una correlazione ecologica, poiché l'unità d'analisi non è individuale ma a livello di gruppo di persone, i residenti di uno stato. La correlazione suggerisce un'associazione positiva tra i nati all'estero e i non analfabeti: è più probabile che i nati all'estero siano non analfabeti rispetto ai nativi. In realtà, l'associazione è negativa. La correlazione calcolata a livello individuale

Tabella 2.1. Notazione per la sezione i -esima. Nelle colonne sono rappresentate le decisioni di voto per il partito A o B al tempo $t + 1$ e nelle righe le decisioni di voto per il partito A o B al tempo t .

	<i>Partito A</i> _{$t+1$}	<i>Partito B</i> _{$t+1$}	
<i>Partito A</i> _{t}	π_a^i	$1 - \pi_a^i$	X'_i
<i>Partito B</i> _{t}	π_b^i	$1 - \pi_b^i$	$1 - X'_i$
	Y'_i	$1 - Y'_i$	1

è infatti di -0.11 . La correlazione ecologica da una errata informazione. Il segno della correlazione è positivo poiché i nati all'estero tendono a vivere negli stati in cui i nativi sono relativamente colti.

Esponiamo alcuni dei metodi citati in precedenza discutendo un caso semplice. Il problema base ha tre variabili osservate Y_i , X_i e N_i e due quantità di interesse non osservate π_a^i e π_b^i per ognuna delle i per ($i = 1, \dots, p$) sezioni elettorali. Denotiamo con:

- X_i il numero di persone che al tempo t votano per il partito A e $N_i - X_i$ il numero di votanti al tempo t per il partito B;
- Y_i il numero di persone che al tempo $t + 1$ votano per il partito A e $N_i - Y_i$ il numero di votanti al tempo $t + 1$ per il partito B;
- N_i rappresenta il numero di persone votanti.

Si sta supponendo un sistema chiuso, ovvero che le somme dei marginali di riga e di colonna siano uguali. Dividendo le due quantità osservate Y_i , X_i per N_i si ricavano le proporzioni marginali di colonna Y'_i , $1 - Y'_i$ e di riga X'_i , $1 - X'_i$. Per una migliore descrizione costruiamo la Tabella 2.1. In questa tabella osserviamo per ogni sezione elettorale i (per $i = 1, \dots, p$) la proporzione di persone che al tempo t votano per il partito A denotato con X'_i e che votano per il partito B chiamato $1 - X'_i$;

e, inoltre, la proporzione di persone che al tempo $t + 1$ votano per il partito A che è Y'_i e per il partito B che è $1 - Y'_i$. Le quantità di interesse che rimangono non osservate a causa del segreto dell'urna sono il numero di elettori che al tempo t avevano votato per il partito A e al tempo $t + 1$ votano per lo stesso partito A ovvero π_a^i ; e, inoltre, il numero di elettori che al tempo t avevano votato per il partito B e al tempo $t + 1$ votano per il partito A ovvero π^i . Lo scopo è di stimare le quantità di interesse π_a^i e π_b^i dalle quantità osservate X'_i e Y'_i . Si possono anche calcolare le quantità:

$$X_a = \frac{\sum_i^p N_i X'_i \pi_b^i}{\sum_i^p N_i X'_i} \quad (2.1)$$

$$X_b = \frac{\sum_i^p N_i (1 - X'_i) \pi_b^i}{\sum_i^p N_i (1 - X'_i)} \quad (2.2)$$

che rappresentano medie, pesate per tutte le sezioni elettorali con il totale dei votanti, dei voti al partito A (2.1) e al partito B (2.2) in tutta l'area considerata.

2.1. Estrazione dell'informazione deterministica: il metodo dei bounds

Lo scopo del *metodo dei bounds* è di estrarre informazione deterministica sulle quantità d'interesse. L'intuizione che sta alla base di questa tecnica è semplice. Esaminiamo il metodo con un esempio. Riferendoci alla Tabella2 .1 possiamo calcolare un intervallo di valori possibili per i parametri d'interesse come segue:

$$\pi_a^i \in [\max(0, \frac{Y'_i - (1 - X'_i)}{X'_i}), \min(\frac{Y'_i}{X'_i}, 1)] \quad (2.3)$$