

Università della Valle d'Aosta - Univeristé de la Vallée d'Aoste

Dipartimento di Scienze Economiche e Politiche

Scienze dell'Economia e della Gestione Aziendale

Anno accademico: 2019/2020

Analisi delle preferenze elettorali attraverso un multinomial mixed logit model bayesiano

Relatore

Chiar.ma Prof.sa Consuelo R. Nava

Correlatore

Chiar.mo Prof. Massimo Angelo Zanetti

Candidato

Zardo Federico

17 C05 573

28-10-2020

*“Supremo è il valore della tenacia: chi fa della costanza la sua parola d’ordine, taglierà
via via tutti i traghetti. Il fallimento è incompatibile con la tenacia, non riuscirà mai
a tenerle testa.”*

– Napoleon Hill

Ringraziamenti

Alla Professoressa Nava, per avermi fatto appassionare alla statistica;

Al Professor Zanetti, per gli utili consigli e per avermi fornito i dati;

Ai miei genitori e ai miei fratelli, che mi hanno sempre sostenuto;

Ai miei nonni, per averci creduto sin dall'inizio e avermi supportato fino alla fine;

Al resto della mia famiglia;

A Lucrezia, che mi ha dato la forza di ricominciare;

A Enrico, per la sua disponibilità e per essersi dimostrato un vero amico;

*A Roberto, compagno di risate e di battaglie, che col suo spirito ha reso più leggeri
questi anni;*

A Veronica e Giorgia, per avermi sopportato durante questo percorso;

A David, per essere stato sempre presente nel momento del bisogno.

Indice

1	Introduzione	6
2	Le preferenze elettorali nelle democrazie avanzate europee	8
2.1	Evoluzione post-bellica delle principali democrazie europee	9
2.2	L'analisi delle preferenze elettorali in uno spazio bidimensionale	14
3	Inferenza bayesiana	31
3.1	Teorema di Bayes	31
3.2	Approccio inferenziale bayesiano	36
3.3	Tecniche di supporto all'inferenza bayesiana	55
3.3.1	Markov Chain	55
3.3.2	Simulazione Monte Carlo	62
3.3.3	Markov Chain Monte Carlo	66
3.3.3.1	Algoritmo di Metropolis-Hastings	70
3.3.3.2	Algoritmo di Gibbs	72
4	Modello logit e sue estensioni	75
4.1	Modello logit binomiale	75
4.2	Modello logit multinomiale	80
4.3	Modello logit multinomiale misto in chiave bayesina per modelli a scelta discreta	83
5	Applicazione del modello logistico multinomiale bayesiano alle preferenze elettorali dei cittadini europei	90

6	Conclusioni	100
	Bibliografia	106
A	Appendice	107
A.1	Fondamenti della probabilità	107
A.1.1	Impostazione assiomatica	108
A.1.2	Concetto di probabilità: vari approcci	110
A.1.2.1	Approccio classico	111
A.1.2.2	Approccio frequentista	112
A.1.2.3	Approccio soggettivista	114
B	Appendice	117
B.1	Variabili aleatorie	117
B.1.1	Variabili aleatorie discrete	118
B.1.2	Variabili aleatorie continue	128
C	Appendice	136
C.1	Codice R	136
C.2	Tabelle excel di supporto al codice	148

1. Introduzione

Questo elaborato ha l'obiettivo di identificare i fattori che influenzano le scelte elettorali nelle principali democrazie capitaliste europee post-industriali. A tal fine si sceglie di utilizzare un approccio Bayesiano parametrico applicato a un modello per scelte discrete con effetti misti.

Nel secondo capitolo si analizzerà quali sono le principali evidenze empiriche che emergono dalla letteratura in merito alle scelte elettorali. Al fine di fornire un quadro generale, il capitolo inizia facendo una breve panoramica sull'evoluzione socio-economica delle principali democrazie europee dal dopoguerra a oggi, prestando particolare attenzione alle conseguenze che le crisi degli anni '70 hanno prodotto sul mercato del lavoro. Segue una riflessione sul fatto che le politiche passate influenzano e vincolano fortemente le politiche odierne, e si raggruppano Paesi che in passato hanno implementato politiche simili, suggerendo così categorie di Paesi omogenee in termini di vincoli istituzionali e, quindi, di offerta politica. Si vedrà che dal punto di vista della domanda politica, i cambiamenti nel tessuto economico e sociale delle società europee negli ultimi decenni hanno determinato una frammentazione dei gruppi sociali, causando una maggiore eterogeneità nell'elettorato: essa si è tradotta in diversi bisogni e diverse sensibilità, e quindi in differenti preferenze politiche.

Questa maggiore eterogeneità che caratterizza l'elettorato delle società post-industriali, ha avuto ripercussioni sui metodi di analisi politica, rendendo anacronistici modelli che si basano su spazi politici monodimensionali, in favore di modelli multidimensionali.

Per descrivere e analizzare tali fenomeni si è scelto di utilizzare un modello bidimensionale per l'analisi delle preferenze elettorali: per individuare le dimensioni viene introdotta la *teoria dei cleavages*, una particolare teoria secondo la quale le preferenze elettorali sono l'espressione di prese di posizione su determinati temi ritenuti rilevanti.

Nel terzo capitolo viene, dunque, introdotto l'approccio inferenziale bayesiano, approc-

cio diverso rispetto a quello classico-frequentista: esso considera i parametri come variabili e non come costanti, sfruttando in estrema sintesi la formula di Bayes. Tale approccio utilizza, quindi, possibili conoscenze a priori, ricalcando più fedelmente la realtà. Nella sezione 3.3, vengono approfondite alcune delle tecniche che hanno permesso di applicare l'inferenza bayesiana anche a problemi inferenziali complessi e che ne hanno permesso l'affermazione in campo scientifico.

Nel quarto capitolo si entra nel cuore del modello utilizzato nell'analisi dei dati: il modello logit multinomiale misto in chiave bayesiana. Per introdurre tale strumento, si è ritenuto utile partire dalla sua versione più semplice e introdurre gradualmente elementi di complessità crescente: i primi step seguono la logica classica, mentre nell'ultima sezione si aggiungono le logiche bayesiane, rendendolo adatto a modelli gerarchici a scelta discreta. La decisione di utilizzarlo deriva dal fatto che, per il tipo di analisi che verranno effettuate, aderisce meglio alla realtà e permette di rilassare alcune ipotesi alla base del modello lineare che risulterebbero inadeguate e restrittive.

L'ultimo capitolo consiste nell'applicazione del modello approfondito a livello teorico e ha come obiettivo trattare il tema del voto politico a livello europeo, considerando che nei vari Paesi analizzati non tutti i gruppi europei sono rappresentati, determinando eterogeneità nel set di scelte per gli elettori. Si prenderà a riferimento l'ottava legislatura del Parlamento Europeo. A tal fine, i dati analizzati sono quelli resi disponibili dall'ultima rilevazione dati dell'European Social Survey, risalente al 2018.

L'obiettivo di questo elaborato è duplice: esso non solo offre un'analisi del comportamento elettorale, ma propone anche l'utilizzo di un modello innovativo al fine di correttamente interpretare i dati disponibili. La struttura della tesi è stata articolata per facilitare anche la comprensione delle logiche bayesiane a coloro che si avvicinano per la prima volta a tale metodo.

2. Le preferenze elettorali nelle democrazie avanzate europee

L'obiettivo di questo capitolo è quello di descrivere i principali fattori che nelle democrazie capitaliste avanzate europee influenzano le decisioni di voto degli elettori o, più precisamente, di gruppi di elettori.

Nella prima parte si descriverà brevemente l'evoluzione di queste democrazie dal secondo dopoguerra a oggi e si analizzerà come i conseguenti cambiamenti nel tessuto socio-economico abbiano influito nel plasmare le attuali preferenze elettorali.

Si parlerà quindi del passaggio da analisi politiche basate su modelli dicotomici unidimensionali dello spazio politico, tipici degli anni del “miracolo economico”, a modelli multidimensionali più complessi, i quali modellizzano meglio l'arena politica nell'età post-industriale, prendendo in considerazione almeno due dimensioni ritenute rilevanti nel plasmare le preferenze. Anche se qualche riflessione verrà fatta sui vincoli che i politici hanno nell'implementare i loro programmi, il focus sarà sui *drivers* delle preferenze politiche, e per individuarli ci si riferirà alla teoria dei *cleavages*.

Si vedrà come, seguendo le logiche di questa teoria, gli elettori si divideranno su temi ritenuti socialmente rilevanti: di conseguenza nasceranno nuovi partiti con l'obiettivo di rappresentarne le posizioni a livello istituzionale. Inoltre, secondo questa teoria, i partiti nati su precedenti *cleavages*, i partiti ora considerati “tradizionali”, fanno molta più fatica a prendere posizioni su queste “spaccature sociali”. Si sostiene, infatti, che gli elettori dei partiti “tradizionali” siano fortemente ancorati ai temi storici portati avanti da questi partiti, e un cambio di posizione nello spazio politico da parte di quest'ultimi potrebbe avere costi in termini di coerenza, credibilità, e quindi di voti: se ciò fosse vero, il numero di voti, di seggi, ottenuti dai nuovi partiti, potrebbero essere una buona *proxy* della rilevanza sociale dei temi alla base delle “spaccature sociali”.

2.1 Evoluzione post-bellica delle principali democrazie europee

Per l'analisi delle preferenze elettorali si sono prese come oggetto di studio le principali democrazie europee. Il motivo di tale scelta è, in primo luogo, l'omogeneità, sotto l'aspetto dei rispettivi sistemi economico-sociali in termini di logiche capitaliste, che permette analisi comparative di tali democrazie. Queste analisi sono tuttavia arricchite dall'eterogeneità che questi Paesi presentano sotto il profilo culturale, linguistico e politico, e dalle diverse reazioni di fronte ai cambiamenti imposti dallo sviluppo tecnologico, dalla globalizzazione e dalle crisi internazionali.

Tali Paesi sono stati assoggettati a un processo di integrazione europea che ne ha limitato la sovranità nazionale, sia in termini di politica economica che di politiche sociali: ciò è stato reso possibile tramite l'istituzione di un articolato sistema di *governance* sovranazionale. È dunque interessante osservare come questo fenomeno di denazionalizzazione comune a democrazie con peculiarità e storie diverse abbia influito sulle preferenze politiche e sui sistemi di partito dei singoli Paesi, o di gruppi di Paesi con caratteristiche simili.

Analizzando l'evoluzione che ha portato a questa situazione, si può iniziare osservando il periodo che parte dalla fine della Seconda Guerra Mondiale alle crisi degli anni Settanta. Questo è stato rinominato "età dell'oro", in quanto ha rappresentato un periodo di ricostruzione, sviluppo ed espansione demografica, avviatosi anche grazie al sostegno finanziario degli Stati Uniti: il Piano Marshall prevedeva infatti lo stanziamento di circa 13 miliardi di dollari per un periodo di quattro anni per la ripresa europea [Blanchard et al., 2014]. I tassi di crescita annui in termini di PIL pro-capite segnarono percentuali senza precedenti: prendendo a riferimento le tre maggiori economie dei sei futuri Stati fondatori della Comunità Economica Europea, la Germania registrò tassi di crescita annui del 6.5%, la Francia del 3.5% e l'Italia del 5.3% (trend registrato anche per la maggiore parte delle altre democrazie capitaliste occidentali, tra cui anche quelle che aderiranno successivamente al progetto europeo) ¹.

Questo fenomeno, unito al forte intervento degli stati nell'economia, fu alla base della convergenza economica tra le democrazie appena citate e della riduzione nei divari

¹<https://www.istat.it/60annidieuropa/storia-e-numeri.html>.

reddituali. Valori così elevati nei tassi di crescita furono principalmente alimentati dagli effetti positivi sul mercato del lavoro che la crescita del settore manifatturiero, causata dalla necessità di ricostruzione di ciò che la guerra aveva distrutto, aveva prodotto [Sessant'anni di Europa].

Il settore manifatturiero era ed è composto principalmente da aziende fortemente immobilizzate, strutturalmente rigide, caratterizzate dalla preponderanza dei costi fissi sui costi variabili e in grado, quindi, di produrre elevati margini di contribuzione nei periodi di alta domanda. Tali guadagni permisero una maggiore occupazione, salari più cospicui e investimenti produttivi, fattori alla base degli elevati tassi di crescita registrati fino agli anni settanta. Ciò nonostante, il divario sociale rimase quello tra operai e imprenditori. L'operaio era supportato dai sindacati e dai partiti socialisti, che si battevano perché avesse sempre maggiori diritti, mentre i detentori di capitale cercavano, attraverso la rappresentanza di partiti tipicamente di destra, di preservare per quanto possibile il loro status. Allora, la realtà politica poteva ancora essere descritta da un modello dicotomico del tipo lavoro contro capitale, sinistra contro destra.

Le crisi inaspettate degli anni settanta portarono, tuttavia, a un cambio piuttosto repentino del contesto sociale appena descritto. La combinazione dell'abbandono degli accordi di Bretton Woods da parte degli USA, con conseguenze sulla stabilità monetaria, e delle crisi petrolifere indotte dai Paesi interni all'OPEC, generò un'inattesa quanto brutale crisi economica. Se è vero che le aziende nel settore manifatturiero producono ottimi risultati economici nei periodi di crescita economica, nei periodi di crisi e di contrazione della domanda esse sono altrettanto sensibili e rischiano perdite cospicue: la rigidità nella loro struttura non permette un ridimensionamento dei costi in risposta a una perdita di ricavi e ciò si traduce in forti crisi aziendali che possono portare anche al fallimento [Anthony et al., 2011]. Quello che infatti accadde alle aziende fu un calo improvviso della domanda e un aumento dei costi energetici.

A livello macroeconomico, la situazione si cristallizzò in un periodo di stagflazione, ovvero un periodo caratterizzato da una crescita pressoché nulla e da inflazione crescente. I minori profitti e la riduzione delle attività produttive derivanti dalla minore domanda, causata dalla combinazione di instabilità monetaria e crisi petrolifere, influirono sulla crescita, mentre l'aumento generalizzato dei costi energetici ebbe ripercussioni dirette sull'inflazione [Blanchard et al., 2014].

La ricetta usata per far fronte al periodo di crisi, fu principalmente di stampo neo-liberista, abbandonando quindi le politiche keynesiane che caratterizzarono gli anni del miracolo economico. Il modello economico preso a riferimento fu quello Thatcher-Reagan, caratterizzato da minor intervento statale in favore del libero mercato e delle privatizzazioni, anche se con alcune eccezioni: alcune democrazie, principalmente quelle scandinave e quella francese, puntarono infatti sul settore pubblico.

Anche il progresso tecnologico ricoprì un ruolo fondamentale: favorì la globalizzazione sia in termini produttivi che di mobilità del lavoro, abbattendo notevolmente le barriere spaziali e temporali, le quali ostacolavano sia il commercio internazionale che la mobilità del lavoro (sia a livello nazionale che transnazionale). Tale progresso fu anche alla base di un processo di terziarizzazione², in quanto le scoperte tecnologiche favorirono tecniche produttive automatizzate che ridussero, e in alcuni casi rimpiazzarono, tutte quelle professioni tayloristiche tipiche del settore manifatturiero. La conseguenza fu una transizione occupazionale dal settore secondario al terziario, cambiando così le dinamiche del mercato del lavoro.

Questo cambiamento spiega una buona parte dei bassi tassi di crescita che le principali democrazie europee registrarono nei decenni successivi: il settore terziario, contrariamente a quello secondario, è composto generalmente da aziende con strutture di costi flessibili, poco immobilizzate, con una prevalenza di costi variabili e, conseguentemente, meno redditizie per definizione, ma anche meno rischiose. Ciò quindi portò, in linea generale, a tassi di crescita più contenuti ma meno volatili. L'attuazione delle politiche neoliberiste volte a contenere l'inflazione invertì, tuttavia, il trend di convergenza economica tra le democrazie europee, e i divari tra i redditi ricominciarono a crescere.

Questi fenomeni ebbero importanti implicazioni politiche e istituzionali. Il sostegno al *welfare* iniziava a essere visto con diffidenza dall'opinione pubblica, in quanto le aspettative sui tassi di crescita non erano delle più rosee: la percezione sulla sostenibilità dell'indebitamento per il finanziamento dello stato sociale iniziò a vacillare e anche le proposte politiche in tal senso persero consensi. Inoltre, la transizione verso un'economia più incentrata sul settore terziario portò a una maggiore eterogeneità nelle professioni presenti sul mercato del lavoro e generò un aumento notevole della domanda di lavoratori altamente qualificati.

²Per terziarizzazione si intende il passaggio da un'economia basata sul settore secondario a un'economia basata sul settore terziario.

Questo fenomeno peggiorò le prospettive di lavoro di operai e impiegati poco qualificati, dei piccoli lavoratori indipendenti e dei proprietari di piccoli negozi, ovvero di tutte quelle categorie alla base della struttura sociale. Si osservò un crescente disinteresse delle associazioni lavorative, principalmente quelle sindacali, che in precedenza tutelavano gli interessi di questi gruppi sociali, lasciando la rappresentanza di queste categorie a partiti emergenti, generalmente più radicali e nati sul malcontento della popolazione.

Il processo di terziarizzazione, ha dunque portato a una riduzione del numero degli operai, soprattutto nel settore industriale, a un'espansione degli impiegati poco qualificati nel terzo settore, ma anche a un'importante crescita dei lavoratori altamente qualificati, sia nel settore pubblico che nel privato.

Una precisazione anche per quanto riguarda il lavoro dipendente è doverosa: il declino del potere sindacale e la de-standardizzazione del mercato del lavoro inasprirono le differenze contrattuali dei lavoratori. Si creò una spaccatura tra il gruppo dei lavoratori a contratto indeterminato e il gruppo di quelli a contratto determinato o senza contratto: il primo, composto da dipendenti altamente tutelati e protetti sul mercato del lavoro; il secondo, composto da dipendenti poco protetti, a rischio disoccupazione e con benefit ridotti, se non nulli. I primi erano tipicamente impiegati nel settore pubblico, mentre gli altri nel privato.

Tutte queste nuove dinamiche nel mercato del lavoro hanno quindi portato a una maggiore eterogeneità occupazionale rispetto a pochi anni prima, generando nuovi gruppi sociali differenti in termini di istruzione, di retribuzioni e di sensibilità politiche sui temi riguardanti la globalizzazione, l'immigrazione, l'intervento dello Stato nell'economia e le politiche economiche e sociali.

Tali gruppi sociali potrebbero essere così suddivisi [Beramendi et al., 2013]:

- Impiegati e dipendenti poco qualificati nel settore secondario e nel settore terziario. In questa categoria, la distinzione tra l'inquadramento contrattuale è significativa: a livello di shock sul mercato del lavoro, l'esposizione dei lavoratori con contratti a tempo indeterminato è completamente diversa rispetto all'esposizione dei lavoratori precari e ciò si riflette sulla sensibilità politica rispetto a determinati temi e quindi sulle preferenze elettorali.
- Piccola borghesia: gruppo composto da soggetti con bassi livelli di istruzione, ma

con patrimoni e redditi relativamente elevati e tipicamente composto da lavoratori autonomi e piccoli imprenditori.

- Professionisti socioculturali: gruppo composto da professionisti con elevati livelli di istruzione e reddito, impiegati principalmente in organizzazioni statali e *no profit*, specialmente in ambito sociale e culturale. Data la struttura gerarchica tipicamente piatta di queste organizzazioni, sono individui con ampie autonomie lavorative e, visto il contesto in cui lavorano, sono spesso a contatto con clienti di ogni classe sociale e nazionalità. Rientrano in questa categoria, per esempio, i professori e i medici.
- Manager, professionisti della finanza, esperti tecnici: gruppo composto da individui con un'istruzione elevata, altamente qualificati. Questi professionisti presentano generalmente elevati redditi e patrimoni consistenti. In questo gruppo è facile trovare ingegneri, investitori, manager e analisti finanziari.

L'analisi delle preferenze e delle caratteristiche di questi gruppi verrà affrontata nella prossima sezione.

Un altro aspetto che ha diviso l'opinione pubblica comunitaria, è stato il processo di crescente integrazione europea che, passando dall'istituzione dell'Unione Europea tramite il Trattato di Maastricht nel '92, ha avuto la sua massima espressione con l'unione monetaria agli inizi del nuovo millennio. Quest'unione monetaria e valutaria, al fine di perseguire obiettivi di contenimento dell'inflazione, ha tolto uno dei due strumenti di politica economica a disposizione dei singoli Stati, ovvero la politica monetaria, riducendo così gli strumenti a disposizione dei governi per portare avanti i rispettivi programmi politici. Anche dal punto di vista giuridico e legislativo questo importante processo di integrazione ha limitato notevolmente la sovranità nazionale dei Paesi membri [Blanchard et al., 2014]. Globalizzazione e denazionalizzazione sono alla base di una nuova dimensione politica che contrappone l'universalismo al particolarismo all'interno dell'ordine sociale e che verrà approfondita nella prossima sezione, in quanto ricoprirà un ruolo principale nei nuovi modelli multidimensionali per le analisi delle preferenze elettorali.

2.2 L'analisi delle preferenze elettorali in uno spazio bidimensionale

Come anticipato, i cambiamenti che hanno visto come protagoniste le principali democrazie capitaliste negli ultimi quarant'anni hanno reso la competizione elettorale molto più complessa. Di conseguenza, i modelli dicotomici che ben rappresentavano la realtà negli anni antecedenti le crisi energetiche, non sono più rappresentativi. Da allora, complici anche tutti i cambiamenti causati dalla tecnologia, i modelli utilizzati per rappresentare lo spazio politico sono diventati più complessi e i fattori da tenere in considerazione per analisi sono aumentati notevolmente.

Tali modelli partono dal presupposto che le preferenze elettorali e le strategie di partito derivino dal fatto che i partiti cerchino di vincere le elezioni e di rimanere all'interno delle istituzioni per più tempo possibile. Inoltre, poggiano le basi sulla presunzione che i politici cerchino di perseguire un determinato programma politico, e non che le loro azioni siano governate solamente da interessi egoistici individuali. Da ciò emerge la caratteristica tipica di operazioni volte a modellizzare la realtà: esse ne sono solo discrete approssimazioni e come tali devono essere interpretate. La certezza di risultato è impossibile da perseguire e questi modelli servono solamente a delineare tendenze più o meno significative.

Ciò detto, essendo le democrazie prese in analisi multipartitiche e le società più diversificate rispetto a un tempo, i partiti con gli obiettivi appena discussi devono spesso ridimensionare le ideologie per favorire la sostenibilità elettorale, fondata su coalizioni e compromessi e sull'adeguamento alle preferenze degli elettori; quest'ultimo aspetto vedremo però esser fortemente controverso in letteratura, a seconda di come gli esperti interpretino la flessibilità dei partiti.

Per l'analisi politica, in questo elaborato ci si concentrerà maggiormente sui fattori che determinano le preferenze elettorali; è tuttavia opportuno fare una piccola riflessione in merito alle eredità politiche dei principali Paesi europei.

I fattori che influenzano i governi attuali nell'attuare le politiche in programma sono, principalmente, le dinamiche sociali e la salute economica dei rispettivi Stati, ovvero fattori tendenzialmente ereditati da politiche passate.

La maggior parte dei Paesi che nel passato hanno attuato politiche sopra le loro possibilità economiche presenta oggi un forte stock di debito. Ciò comporta un minor margine economico per perseguire politiche attuali, mentre ulteriori manovre a debito possono provocare tensioni sui mercati e portare, come è successo, le istituzioni sovranazionali (in questo caso l'Unione Europea) a imporre dei vincoli sulla quantità di *deficit* che gli Stati sovrani possono prevedere nella legge di bilancio: le reazioni dei mercati e l'intransigenza dell'UE sono ancora più marcate quando tale spesa a debito è destinata al sostegno di politiche di breve periodo tipicamente finalizzate a generare consenso e non crescita di medio-lungo termine.

La crescita è lo strumento più potente per la sostenibilità e per la riduzione del debito passato e, dal punto di vista macroeconomico, i vincoli sul deficit contribuiscono ad aumentarla, al fine di generare benefici sulla salute dell'economia futura. I politici, però, dovendo rispondere alle esigenze di breve periodo dei cittadini e avendo fini elettorali, tendono a sfiorare questi vincoli di medio-lungo periodo e a concepirli come un'intermissione esterna e una riduzione della sovranità nazionale. Le politiche di medio-lungo termine, per la natura stessa della democrazia e per la scarsa consapevolezza popolare, non generano voti, mentre politiche di breve rispondono ai bisogni dei cittadini, riflettendosi dunque nella loro scelta di voto [Blanchard et al., 2014].

Non si intende però far passare il messaggio che politiche di breve siano il male assoluto: esse infatti sono l'espressione democratica delle esigenze e delle necessità del popolo sovrano e, dal punto di vista sociale, è giusto che vengano implementate. Si pensi alla crisi economica causata dal COVID-19: non sostenere le categorie più colpite con politiche mirate potrebbe creare danni sociali che potrebbero trasformarsi in danni economici più seri se confrontati a quelli derivanti da un ulteriore indebitamento per contenere la crisi. Si potrebbe assistere a un aumento della criminalità, del lavoro in nero, della sfiducia nello Stato, dell'emigrazione, ovvero tutti fattori che potrebbero generare grossi problemi per la tenuta di una democrazia.

Una politica illuminata, a parer di chi scrive, dovrebbe tuttavia trovare l'equilibrio tra le scelte di breve e di medio-lungo periodo, puntando su queste ultime nei periodi di benessere e utilizzando le prime nei casi di forti shock esogeni e di forti crisi. Quello che però si è osservato in alcuni Paesi europei, tra cui l'Italia, è stato un abuso di politiche di breve periodo, principalmente con lo scopo di intercettare consensi e vincere le elezioni, anche nei periodi in cui politiche di più lungo periodo sarebbero state sostenibili e

avrebbero generato maggiori benefici per l'economia futura.

Tuttavia, non tutte le democrazie europee hanno avuto una classe politica orientata al breve periodo e questo rappresenta un primo fattore alla base dell'eterogeneità tra i Paesi presi in considerazione nella parte introduttiva di questo capitolo. Queste democrazie differiscono, quindi, in termini di eredità politica e ciò può essere catturato da una dimensione che contrappone le scelte che sono state fatte negli anni dai diversi governi nazionali. Per modellizzare nel modo più semplice questa dimensione e per essere in linea con [Beramendi et al., 2013], alle politiche di breve periodo si associano le azioni volte a sostenere i consumi e i trasferimenti, mentre alle politiche di medio-lungo periodo quelle volte a incentivare gli investimenti, sia diretti (per esempio investimenti in ricerca e sviluppo e in infrastrutture) che indiretti, attraverso incentivi, spesso fiscali. C'è un'altra dimensione che differenzia le democrazie prese in considerazione, ovvero quanto l'intervento dello Stato abbia pesato nell'economia: in alcune Nazioni è stato molto marcato, mentre in altre è stato minore, in linea con le teorie liberiste che sostenevano e sostengono che l'intervento dello Stato nell'economia provochi più danni rispetto al lasciare che i mercati si autoregolino [Blanchard et al., 2014].

Dalla combinazione di queste due dimensioni è possibile creare categorie di Paesi con caratteristiche simili: riprendendo la terminologia utilizzata in [Beramendi et al., 2013], si avranno: Paesi cattura, Paesi caratterizzati da competitività, Paesi che puntano all'uguaglianza e Paesi che puntano allo Status.

Ogni categoria, come ben rappresentato in figura 2.1, è caratterizzata dall'esser vincolata da una certa combinazione delle due dimensioni, combinazione frutto principalmente delle scelte politiche passate.

I Paesi che puntano sulla competitività sono caratterizzati da un basso intervento dello Stato e da politiche che prediligono gli investimenti sui consumi: è la combinazione che dal punto di vista macroeconomico risulta essere la più prudente e la più sostenibile, in quanto orientata al medio-lungo periodo e che comporta meno interferenze nelle dinamiche di mercato. Tali interferenze, anche se attuate con i miglior propositi, tendono in genere a distorcere l'economia producendo effetti incerti e a volte controproducenti [Blanchard et al., 2014]. I Paesi che rientrano in questa categoria sono quelli anglosassoni.

I Paesi che puntano sull'uguaglianza, sono anch'essi più inclini agli investimenti. Tut-

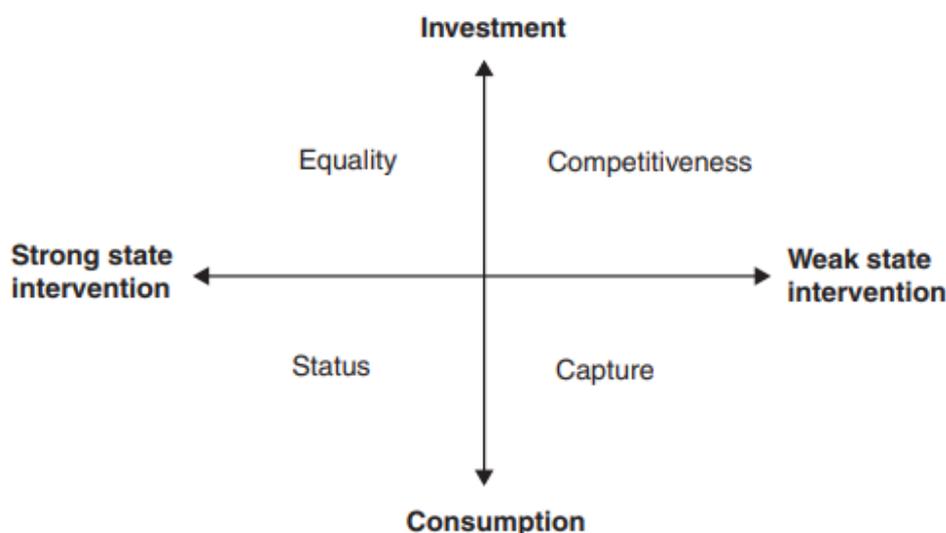


Figura 2.1: Vincoli istituzionali. Fonte: [Beramendi et al., 2013].

tavia, a differenza di quelli descritti in precedenza, essi presentano un forte intervento dello stato nell'economia. Questa combinazione, nel corso degli anni, si è dimostrata essere una scelta vincente in termini di un'equa distribuzione della ricchezza, con ricadute positive sulla società e sul benessere generale [Beramendi et al., 2013]. Ciò è un po' in contraddizione con le teorie economiche più liberiste, che vedono nell'intervento statale un blocco alla crescita e allo sviluppo. Potrebbe essere anche la storia di questi Paesi, in termini di qualità della classe politica passata, ad aver influito sulla situazione attuale [Blanchard et al., 2014]. In questi Paesi, principalmente scandinavi, si registrano infatti i più bassi tassi di corruzione [Cucciniello et al., 2020], e ciò è significativo, in quanto, come spiegato in [Blanchard et al., 2014], è una delle dinamiche sociali che rallenta lo sviluppo e la crescita: essa ostacola la meritocrazia, produce inefficienze e produce distorsioni nell'impiego di risorse pubbliche, indirizzandole al fine di soddisfare interessi soggettivi piuttosto che, come dovrebbe essere, gli interessi della collettività. Vi è, inoltre, il blocco delle democrazie dell'Europa continentale, dei Paesi "status", caratterizzato da una forte predisposizione ai consumi e un massiccio intervento dello Stato. Quest'ultimo si esprime in politiche volte a incentivare il consumo, tramite una complessa legislazione a tutela del consumatore, una regolamentazione in termini di concorrenza e politiche economiche e fiscali in tal senso. Infine, vi sono gli Stati "cattura", caratterizzati da una forte propensione al consu-

mo e da una bassa capacità di intervento statale. In queste Nazioni, principalmente mediterranee, la politica pubblica opera su base selettiva per compensare gli squilibri che le dinamiche di mercato producono, rispondendo alle richieste di brevissimo periodo di determinati gruppi sociali. Ciò comporta politiche di breve periodo di stampo assistenzialista che non arrivano alle radici dei problemi, ma cercano di tamponarne continuamente gli effetti. Le conseguenze di tali politiche, oltre a non risolvere le problematiche strutturali, si traducono in un'elevata probabilità che piccoli, ma ben organizzati gruppi sociali, "catturino" parti dello Stato per portare avanti i propri interessi [Beramendi et al., 2013].

Questa suddivisione, permette di analizzare le preferenze politiche di Paesi con caratteristiche simili e confrontare poi i risultati tra le diverse categorie.

Dato un ampio ventaglio di proposte politiche portate avanti dai diversi partiti, l'obiettivo ora è individuare i fattori che spingono gli elettori a prediligere determinati partiti piuttosto che altri.

Un parallelismo con il marketing potrebbe rendere bene l'idea: si pensi a un'impresa (sistema di partito) che commercializza differenti tipologie di prodotti (partiti) a clienti eterogenei (elettori); interessante sarebbe analizzare i fattori che influiscono sulle preferenze dei consumatori nel prediligere un prodotto rispetto a un altro, in quanto permetterebbe politiche commerciali e di marketing mirate a intercettare al meglio le dinamiche d'acquisto (preferenze politiche).

Per analizzare le preferenze elettorali bisogna prendere a riferimento i gruppi sociali citati in precedenza, ovvero quelli creatisi nell'epoca post-industriale: si ricordi che essi differiscono principalmente in termini di istruzione, di salari, di protezioni sul mercato del lavoro e di prestigio nella piramide sociale. Molti studi di questo genere tendono ad analizzare gruppi di persone in base all'occupazione ricoperta durante l'età lavorativa e, anche se può risultare una buona approssimazione della realtà, questo metodo di analisi potrebbe risultare impreciso. Intere categorie vengono escluse dalle analisi o vengono fatte delle ipotesi che potrebbero risultare generalizzazioni poco rappresentative della realtà.

Si pensi ai pensionati, ai disoccupati, agli inoccupati e agli studenti con diritto di voto. Si tratta di categorie che non rientrano nell'analisi, in quanto non associabili a nessun gruppo sociale. Alcuni autori, per fare in modo di non trascurare queste categorie,

procedono con il generalizzare che, per esempio, gli studenti rientrano nel gruppo sociale dei genitori o del genitore con il lavoro dominante, oppure associano ai pensionati le caratteristiche del gruppo sociale di appartenenza quando erano in età lavorativa.

In molti casi, tuttavia, queste generalizzazioni sono poco rappresentative della realtà: esistono svariati casi di figli con idee politiche distanti da quelle dei genitori, a volte addirittura opposte, oppure di pensionati che, visto il loro cambio di status, cambiano bisogni e visione politica rispetto ai tempi in cui erano in età lavorativa. Bisognerebbe dunque fare degli studi appositi per queste categorie e, ciò, comporterebbe analisi più complesse e più impegnative in termini di tempo e di costi. Alla ricerca di un'impossibile rappresentazione perfetta della realtà, si preferisce dunque un modello meno accurato ma che richieda meno risorse per essere implementato.

Partendo dunque da questa premessa, si cercheranno di mappare le posizioni politiche dei gruppi sociali visiti in precedenza nello spazio delle preferenze politiche. Lo spazio delle preferenze politiche in età post-industriale, non può più essere banalizzato in uno spazio unidimensionale, ma deve essere articolato in almeno due dimensioni.

Si ritiene opportuno, in linea con [Beramendi et al., 2013], mappare le preferenze su uno spazio che combina una dimensione “stato-mercato” con una dimensione “particolarismo-universalismo”. Queste dimensioni sono l'espressione di quanto esposto nella *teoria dei cleavages*, ovvero una prospettiva basata sul concetto di fratture sociali talmente significative da dividere i membri di una comunità in gruppi: le posizioni di questi gruppi si trasformano in specifiche preferenze politiche che a loro volta si trasformano in voti e quindi in seggi [Capano et al., 2014].

Questa teoria parte dal presupposto che le crisi alla base dei *cleavages*, delle fratture sociali, in questo caso le due dimensioni citate in precedenza, sono l'espressione delle pressioni che si sono accumulate nel tempo rispetto a certi temi. I *cleavages*, secondo questa logica, vedono come protagonisti forze sociali esogene ai sistemi di partito, ovvero gli elettori, i quali, attraverso il loro comportamento elettorale, ne influenzano la composizione e i cambiamenti. Non si attribuisce alcuna importanza ai partiti già presenti nelle istituzioni nel plasmare i sistemi di partito, in quanto sono considerati strutturalmente rigidi, inflessibili: la natura dei cambiamenti nei sistemi di partito viene considerata prettamente esogena e si coglie nella forma di partiti emergenti, espressione delle variazioni nel comportamento elettorale.

Il cambiamento dei sistemi di partito non è, secondo questa teoria, associabile a un processo di riallineamento, ovvero a un processo che nel corso del tempo riallinea le posizioni dei partiti alle preferenze degli elettori, ma è governato dalla discontinuità derivante dall'ingresso di nuove forze politiche a seguito di nuovi *cleavages*. Ciò che determina il dinamismo dei sistemi di partito sono, dunque, i partiti emergenti, espressione delle preferenze elettorali, piuttosto che cambi di posizione dei partiti storici.

Infatti, in sistemi di partito con alte barriere all'entrata come quelli inglesi e francesi, nuove fratture sociali potrebbero generare aspre frizioni tra i partiti, mentre in sistemi con basse barriere all'entrata queste tensioni si scaricano nella forma di nuovi partiti d'opposizione che entrano in competizione con i partiti storici creatisi a loro volta su *cleavages* precedenti³ [Hooghe and Marks, 2018].

Le due dimensioni dello spazio delle alternative politiche sottostanno alle logiche dei *cleavages*. La prima dimensione vede infatti contrapposte preferenze per l'intervento dello Stato nell'economia a preferenze più liberiste. Soprattutto per le democrazie europee che hanno aderito all'unione monetaria, non ha più molto senso analizzare questa dimensione dal punto di vista di politiche macroeconomiche: tra i due principali strumenti a disposizione, ovvero la politica fiscale e quella monetaria, la seconda è stata completamente sottratta dalle competenze nazionali e la politica fiscale è fortemente condizionata da vincoli imposti dall'Unione. Ha senso, invece, pensare a questa dimensione in termini di regolamentazione del mercato del lavoro e di politiche di *welfare*, in quanto ancora sotto la competenza dei governi nazionali e alla base di divergenze significative sia nell'elettorato che nelle politiche di partito.

Le preferenze in questa dimensione vedono dunque contrapposti i soggetti che appoggiano un maggior intervento dello Stato in termini di redistribuzione delle risorse e di maggior intervento nel mercato del lavoro a favore dei lavoratori più deboli, tramite investimenti sociali, e i soggetti più inclini a lasciar che il mercato si autoregoli. Se in prima battuta potrebbe sembrare che il reddito e il livello d'istruzione siano le variabili cardine attorno alle quali si polarizzano le preferenze, con i soggetti che presentano alti livelli di istruzione e di reddito a favore di posizioni più liberali, mentre coloro poco istruiti e con bassi livelli di reddito a favore dell'intervento statale, analisi sociologiche hanno dimostrato che non è così: a favore dell'intervento dello Stato vi sono, oltre ai

³Per esempio i *cleavages* di classe e di religione, i quali hanno avuto importanza centrale nel formare i partiti che oggi vengono considerati "storici".

lavoratori poco qualificati, anche i professionisti socioculturali, mentre a favore del libero mercato vi sono sia i manager e professionisti con elevati livelli di istruzione e di reddito, sia la piccola borghesia, la quale è contraria a ogni forma di intervento statale che non sia nella forma di sussidi alle imprese [Häusermann and Kriesi, 2015].

La seconda importante dimensione che negli ultimi anni è alla base dei maggiori conflitti nell'opinione pubblica, è quella che vede contrapposti universalismo e particolarismo. Questo *cleavage* è legato ai processi di globalizzazione che, soprattutto a partire degli anni Novanta, hanno colpito le principali democrazie occidentali [Beramendi et al., 2013].

La globalizzazione, in quanto processo planetario di integrazione economica e culturale, impatta principalmente sulle dinamiche del commercio, degli investimenti, dell'immigrazione e della mobilità del lavoro, tutti aspetti con ricadute importanti nel mercato del lavoro e conseguentemente sulle preferenze elettorali.

Questo processo ha creato nuove importanti divisioni sociali tra chi sostiene posizioni a favore di un mondo più connesso e tollerante, ovvero i cittadini con una concezione più universalista, e i cittadini contrari alla globalizzazione, con una visione del modo più particolarista. Questa divisione ha messo in secondo piano i *cleavages* precedenti (principalmente le divisioni di classe sociale, di religione) e messo in difficoltà i partiti nati su di essi, i partiti "tradizionali". Infatti, da questa nuova divisione, sono nati partiti che nel tempo hanno sottratto consensi a tali partiti, il che si è cristallizzato, a livello di Parlamento Europeo, in una diminuzione dei relativi seggi. La caratteristica di questi nuovi partiti è quella di sostenere posizioni radicali, non moderate, rispetto alla divisione oggetto di conflitto, in quanto consapevoli del fatto che generano maggiori consensi. Queste posizioni radicali, si dividono in due fronti sociali: il polo TAN (Tradition/Authority/Nationalism) e il polo GAL (Green/Alternative/Libertarian), sostenuti a livello politico rispettivamente dai partiti di destra radicale e dai partiti verdi, ovvero dai partiti emergenti [Hooghe and Marks, 2018] [Dancygier and Walter, 2015].

Per cercare di capire come si schierano i cittadini rispetto a questi due poli, è fondamentale un approfondimento degli effetti della globalizzazione per individuare i gruppi sociali i cui interessi vengono lesi e quelli che invece ci guadagnano da tale processo.

La globalizzazione in termini di commercio internazionale e di internazionalizzazione della produzione, pur generando benefici ai consumatori in termini di prezzi, di varietà di prodotti e, più in generale, in termini di ripercussioni positive sulla crescita

dell'economia, è alla base di rilevanti scompensi distributivi.

Per esempio, nei settori esposti alla competizione internazionale, le imprese più produttive traggono vantaggi dall'apertura ai mercati internazionali. Esse possono infatti vendere i loro prodotti sui mercati esteri oltre che nei mercati di appartenenza, con l'obiettivo di generare ricavi maggiori, e delocalizzare facilmente funzioni aziendali, con l'obiettivo di abbattere i costi: la combinazione di queste due strategie, se implementate correttamente, è finalizzata a generare maggiori profitti. D'altro canto, le imprese meno produttive sono più inclini a soffrire le pressioni di una competizione estesa a livello globale e rischiano di uscire sconfitte dalle sfide che essa comporta.

Questo fenomeno ha ripercussioni importanti anche sul mercato del lavoro: supponendo che all'aumentare del livello di competenze aumenti anche il livello di produttività dei lavoratori, le imprese più produttive saranno inclini ad assumere forza lavoro altamente qualificata e a poterla pagare profumatamente. Questi soggetti, inoltre, beneficeranno dei minori prezzi e della più ampia gamma di prodotti disponibili derivanti dalla globalizzazione.

I lavoratori meno qualificati saranno, quindi, destinati a lavorare nelle imprese meno produttive e meno facoltose, ricevendo generalmente salari minori e diventando più sensibili alla disoccupazione.

Da ciò si deduce che, in merito al commercio internazionale, i lavoratori più qualificati e i proprietari delle imprese più produttive, sosterranno generalmente politiche universaliste, mentre i lavoratori meno qualificati e i proprietari delle imprese meno produttive, saranno inclini a politiche più autarchiche, di stampo particolaristico.

Da questa riflessione emerge già una tendenza interessante: i lavoratori più deboli, una volta sostenuti dai partiti di sinistra, sono oggi principalmente schierati su posizioni particolaristiche, sostenute da partiti di destra radicale, mentre i lavoratori più qualificati, più forti sul mercato del lavoro, sono più inclini a sostenere i Verdi, in quanto sostengono posizioni più universalistiche.

C'è però da fare una precisazione: la tendenza è molto più debole per i lavoratori impiegati in settori non aperti al commercio internazionale, come ad esempio i dipendenti pubblici, in quanto i rispettivi stipendi non dipendono dalle dinamiche commerciali. Questa categoria di lavoratori presenta meno differenze reddituali rispetto ai lavoratori di settori aperti alla competizione internazionale: i lavoratori altamente qualificati

guadagnano un po' meno e quelli poco qualificati un po' di più rispetto alle controparti nei settori aperti alla competizione internazionale, ma soprattutto sono molto più protetti a livello contrattuale. Ciò quindi si riflette in termini di preferenze di voto tendenzialmente più moderate.

Un altro aspetto più che mai attuale e politicamente rilevante è l'immigrazione: questo fenomeno cambia, in parte, la carte in tavola, in quanto influisce sulle preferenze più moderate di quelle categorie che abbiamo ritenuto essere protette dalle dinamiche del commercio internazionale e dalle delocalizzazioni. Si è dimostrato estremamente rilevante come tema in termini politici: i partiti che hanno basato le rispettive campagne elettorali esaltando questo fenomeno, prendendo posizioni estreme su entrambi i fronti, hanno ottenuto sempre più consensi a discapito dei partiti tradizionali, sia di destra che di sinistra, i quali non avevano mai percepito tale fenomeno come rilevante, e ora sono spinti a darne importanza per contenere la perdita di consensi. Capire i motivi di tali dinamiche è dunque fondamentale.

I temi legati all'immigrazione hanno importanti ripercussioni sia sul mercato del lavoro che a livello socio-culturale, soprattutto per le categorie di lavoratori meno qualificati.

L'inserimento di nuovi soggetti nel mercato del lavoro genera competizione, quindi pressioni ribassiste in termini di salari e possibili riduzioni delle protezioni contrattuali. Solitamente, a emigrare sono i soggetti meno qualificati, i quali entrano in competizione con i lavoratori poco qualificati del Paese ospitante, che siano essi protetti o meno in dagli effetti derivanti dall'internazionalizzazione della produzione e del commercio.

I lavoratori altamente qualificati, anche in questo caso, traggono benefici dalla concorrenza, in quanto colpisce maggiormente i lavoratori poco qualificati e i relativi salari: ciò si potrebbe tradurre in prezzi di beni e servizi ancora minori. Tuttavia, c'è una divisione in letteratura rispetto al perché gli individui altamente qualificati siano tendenzialmente favorevoli all'immigrazione: un filone sostiene che alla base di questa tendenza vi siano i motivi prettamente economici appena citati, mentre l'altra linea di pensiero sostiene che sia il livello d'istruzione tipicamente elevato di questi individui ad aumentarne i rispettivi livelli di tolleranza sociale.

La seconda linea di pensiero sembra essere validata dal fatto che gli individui con elevati livelli di istruzione sono tendenzialmente favorevoli anche a un'immigrazione qualificata, anche se potenzialmente potrebbe avere ripercussioni negative sul loro status sia

economico che sociale. Sarebbe interessante capire se l'influenza dell'istruzione su una propensione più universalista in merito all'immigrazione sia frutto di fattori oggettivi, come la conoscenza dei benefici alle imprese che la concorrenza sul mercato del lavoro produce, gli effetti positivi sulla disoccupazione naturale indotti da un minor costo del lavoro e da minori protezioni contrattuali, oppure se è frutto di un'influenza ideologica, magari anche non intenzionale, da parte di chi insegna.

Per quanto riguarda i lavoratori poco qualificati, invece, è condivisa l'idea che essi siano contrari all'entrata di lavoratori stranieri, qualora si tratti di individui poco qualificati, mentre sono più tolleranti se si tratta di individui altamente qualificati. Per questa categoria sembra quindi essere significativa la tesi che mette in primo piano questioni economiche.

Ricapitolando, le categorie che più ci perdono dalla globalizzazione sono i lavoratori poco qualificati e con professioni esposte all'internazionalizzazione della produzione e del commercio: nel caso in cui siano impiegati in aziende produttive, essi rischiano che i datori di lavoro delocalizzino la produzione all'estero, con la conseguenza di perdere il lavoro o di dover accettare di spostarsi nei Paesi in cui viene delocalizzata⁴, nel caso in cui siano impiegati nelle aziende meno produttive che hanno problemi a delocalizzare, rischiano di subire gli effetti negativi che del commercio internazionale. Infine, a livello domestico devono fare i conti con la competizione dei lavoratori stranieri, con conseguenze generalmente negative sui salari e sulle protezioni sul mercato del lavoro⁵. I lavoratori poco qualificati con professioni protette dall'internazionalizzazione della produzione e del commercio, sono tuttavia sensibili all'altro lato della globalizzazione, l'immigrazione. Questa categoria è quindi la plausibile alleata dei lavoratori poco qualificati esposti all'internazionalizzazione della produzione e del commercio, in termini di preferenze politiche più particolaristiche.

Tra i lavoratori altamente qualificati protetti dall'internazionalizzazione dell'economia e quelli che ne sono più esposti, le preferenze politiche potrebbero essere un po' più

⁴Ciò potrebbe avere implicazioni sia morali, in quanto spostarsi in un altro Paese è già complicato di per sé per questioni di lingua e cultura, che economiche, in quanto tendenzialmente le imprese delocalizzano in Paesi con costo del lavoro minori.

⁵Si è specificato generalmente in quanto, in alcuni casi, il lavoro degli stranieri è complementare a quello domestico e ciò potrebbe generare benefici agli stessi lavoratori domestici. Potrebbe anche fungere da scudo nel caso di licenziamenti, in quanto i datori di lavoro potrebbero scegliere di licenziare per primi i lavoratori stranieri.

distanti, in quanto i primi godono solo degli effetti del commercio internazionale e dell'immigrazione, in termini di prezzi di beni o servizi minori, mentre i secondi godono anche di ulteriori benefici dell'internazionalizzazione della produzione.

Tuttavia, anche se in misura diversa, i componenti di queste due categorie tendono a essere più propensi a sostenere partiti con una visione più universalista e più favorevole a ulteriori processi di integrazione socio-economica [Dancygier and Walter, 2015] [Capello, 2004].

Tra questi processi d'integrazione, anche l'integrazione politica gioca un ruolo fondamentale nel contrapporre visioni universaliste e particolaristiche nell'opinione pubblica. In parte, si ricollega a quanto detto in precedenza sul perché, oggi, per i Paesi dell'UE non abbia quasi più senso di trattare la dimensione Stato-mercato in termini di politiche fiscali e monetarie, ma in termini di decisioni che gli Stati possono prendere all'interno di vincoli sovranazionali. Da quanto appena affermato, si denota che il trasferimento dell'autorità politica a istituzioni sovranazionali indebolisce le sovranità dei singoli Paesi: ciò rischia scatenare reazioni nazionaliste, in quanto si è dimostrato in letteratura che l'idea di democrazia è strettamente legata all'idea di Stato-Nazione.

Tuttavia, anche in questo caso, il livello d'istruzione sembra giocare un ruolo fondamentale: i cittadini con maggiori livelli d'istruzione tendono vedere l'indebolimento della sovranità nazionale come un fattore positivo e non come un rischio per la democrazia, confermando così la loro tendenza più universalista, mentre i cittadini meno istruiti si identificano maggiormente nelle rispettive comunità nazionali, ne attribuiscono maggior valore, percependo questo processo come minaccia e prendendo posizioni fortemente particolaristiche avverse alla globalizzazione.

Un ultimo aspetto che si ricollega alle questioni redistributive della dimensione Stato-mercato, è legato a quali soggetti dovrebbero avere il diritto di beneficiare delle risorse pubbliche: i cittadini meno istruiti e più legati a concezioni particolaristiche, ritengono, in genere, che gli aventi diritto all'assistenza sociale dovrebbero esser solamente pochi e meritevoli, solo i contribuenti e, nei casi più estremi, solo i contribuenti residenti. Al contrario, i cittadini più istruiti tendono, in linea generale, a essere favorevoli all'utilizzo dei soldi pubblici anche per aiutare individui non residenti e non contribuenti come, per esempio, gli immigrati clandestini, dimostrando ancora una volta valori universalistici [Häusermann and Kriesi, 2015].

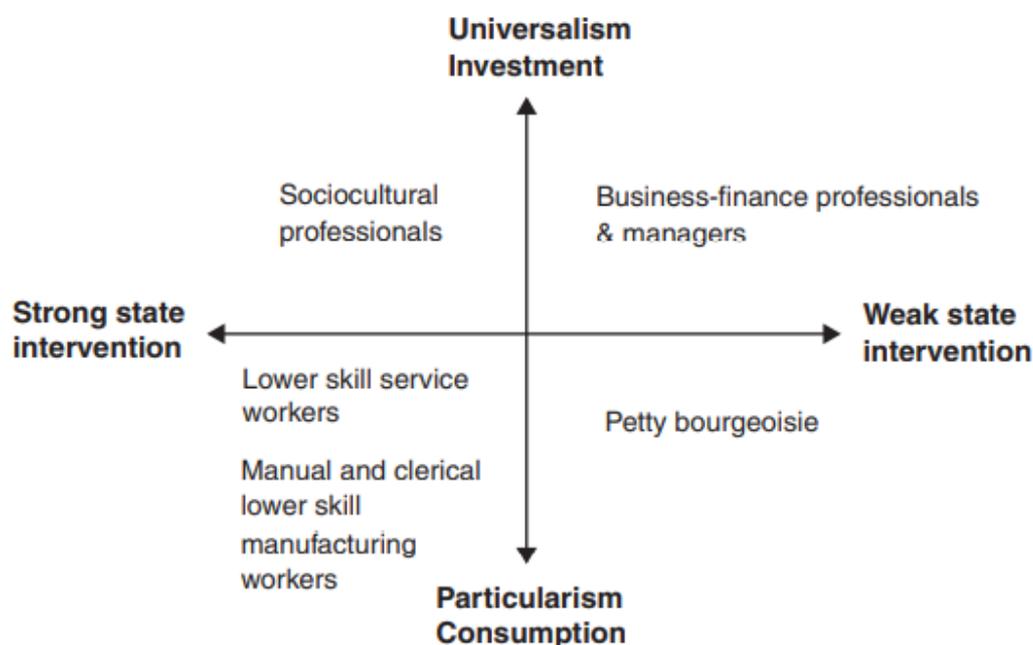


Figura 2.2: Spazio delle preferenze politiche. Fonte: [Beramendi et al., 2013].

Avendo ora chiarito i fattori che generano posizioni contrastanti tra gli elettori in merito alle due dimensioni prese in considerazione per l'analisi delle preferenze, si possono mappare le preferenze per i gruppi sociali descritti in precedenza, ben sintetizzate dalla figura 2.2.

I professionisti socio-culturali, tenderanno a sostenere posizioni politiche universalistiche e a favore di un massiccio intervento dello stato nell'economia.

I manager, i professionisti della finanza, in generale le professioni altamente qualificate, sono anch'essi generalmente favorevoli a posizioni universaliste, ma sono contrari all'intervento dello Stato nell'economia, prediligendo politiche più liberiste.

Queste due categorie, oltre ad avere affinità su una visione universalista, tendono ad avere anche la stessa visione, in termini qualitativi, per quanto riguarda spesa pubblica⁶: entrambe preferiscono che venga utilizzata per politiche a favore degli investimenti, in quanto fortemente contrari a quelle politiche di breve periodo a sostegno dei consumi. Si è specificato che il punto in comune è l'utilizzo della spesa in termini qualitativi,

⁶Sia la propensione a una società più universalista che le posizioni su come dovrebbero essere spesi i soldi pubblici, potrebbero avere la stessa determinante: l'alto livello d'istruzione.

poiché termini quantitativi divergono nelle vedute: i professionisti socio-culturali sono per un forte utilizzo della spesa, mentre gli altri professionisti ne limiterebbero, per quanto possibile, l'utilizzo.

La piccola borghesia, come il gruppo dei manager, sono più inclini a politiche a favore del libero mercato, e sono contrari all'intervento dello Stato nell'economia se non nella forma di sussidi alle imprese. La piccola borghesia diverge però dal gruppo appena citato per quanto riguarda il posizionamento sulla dimensione particolarismo-universalismo: essa infatti prende posizioni fortemente particolaristiche. Si ricorda, al fine di comprendere il perché di questa posizione, che in questa categoria vi sono i lavoratori autonomi e i proprietari di piccole imprese, ovvero quei soggetti che più hanno da perderci dagli effetti della globalizzazione. Anche il loro livello d'istruzione generalmente basso, potrebbe influire su questa posizione, così come sulla loro propensione a sostenere politiche in favore dei consumi⁷.

Infine, vi sono gli operai e i dipendenti poco qualificati: questi soggetti, come i professionisti socio-culturali, sono schierati per un intervento massiccio dello Stato, ma sono tendenzialmente orientati a sostenere politiche di breve periodo orientate a incentivare il consumo come la piccola borghesia.

Se in linea generale affermare che questa categoria prediliga i consumi è sensato, bisogna però precisare che al suo interno vi è una divisione di vedute tra i lavoratori poco qualificati impiegati nel settore manifatturiero e quelli impiegati nel settore dei servizi: gli impiegati nel settore terziario, ricoprendo incarichi lavorativi generalmente più instabili, tendono a prediligere gli investimenti ai consumi, principalmente per lo sviluppo del capitale umano. Il fatto che a prevalere siano però le tendenze dei lavoratori nel settore manifatturiero, è indice a livello politico abbiano ancora maggior peso, siano più rappresentati, i lavoratori del settore secondario.

Questa categoria, condivide con la piccola borghesia le posizioni particolaristiche: i bassi livelli d'istruzione dei componenti di questa categoria e il fatto che siano tra le più sensibili agli effetti negativi della globalizzazione, potrebbero ben spiegare questa presa di posizione.

⁷Ovviamente, oltre al livello d'istruzione, a incidere sulle preferenze della piccola borghesia in merito all'utilizzo della spesa pubblica per incentivare i consumi, potrebbero esserci interessi economici, in quanto politiche a favore dei consumi si traducono generalmente in una maggiore propensione ai consumi da parte dei consumatori e quindi in maggiori ricavi per i commercianti.

Tutti i punti di contatto tra i gruppi sociali che sono stati evidenziati e che sono ben rappresentati in figura 2.3, potrebbero rappresentare per i politici delle occasioni per ideare dei programmi volti a rispondere alle esigenze elettorali di più di un gruppo sociale, al fine di ottenere maggiori consensi.

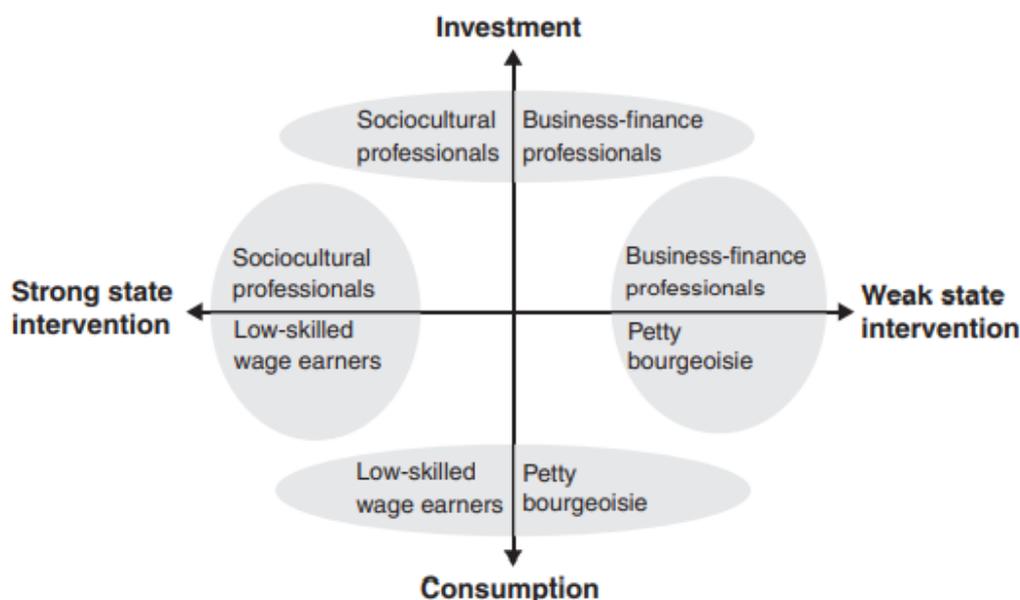


Figura 2.3: Possibili punti di contatto tra gruppi sociali. Fonte: [Beramendi et al., 2013].

A questo punto, è interessante evidenziare come tali posizionamenti si traducano in preferenze: il fatto che gli individui oggetto di studio appartengano a diversi Paesi europei caratterizzati da differenti sistemi di partito, renderebbe le analisi comparative molto complesse. La soluzione è analizzare i gruppi di partito europei, anche chiamati “famiglie europee”, in quanto rappresentano gruppi di partiti politici di diverse nazionalità ma affini a livello politico .

In [Häusermann and Kriesi, 2015], le preferenze elettorali sono state analizzate per le seguenti famiglie europee, ovvero i Verdi, la Sinistra radicale, i Socialdemocratici, i Liberali, i Conservatori cristiani e i populistici della Destra radicale, in base al partito nazionale che, nell’ESS⁸ (2009), gli intervistati hanno dichiarato di votare o di sentire più vicino. Per le analisi, sono state considerate come variabili dipendenti le famiglie

⁸ Acronimo che sta per European Social Survey: si tratta di un sondaggio biennale transazionale finalizzato a indagare le attitudini e i comportamenti dei cittadini europei, utilizzando i più elevati standard

europee e come indipendenti le due scale di preferenza relative alle dimensioni Stato-mercato e universalismo-particolarismo. Inoltre, sono state considerate delle variabili di controllo tra cui l'età, il genere, il reddito, il livello d'istruzione, la classe, l'appartenenza sindacale, la fede nella Chiesa e l'occupazione nel settore pubblico. Inoltre, il modello statistico utilizzato, è stato implementato per blocchi di Paesi con caratteristiche strutturali simili, ovvero analizzando separatamente gli individui appartenenti ai Paesi orientati allo status, alla concorrenza, all'uguaglianza e alla cattura.

Per esempio, prendendo in considerazione il modello per i Paesi orientati allo status, le posizioni sulla scala particolarismo-universalismo sono statisticamente rilevanti e hanno un forte effetto per quanto riguarda la scelta dei partiti emergenti: a posizioni universalistiche più marcate corrispondono maggiori probabilità di votare i Verdi, mentre per posizioni più particolaristiche vi sono maggiori probabilità che il gruppo di partito scelto sia la Destra radicale populista, come sostenuto in precedenza⁹. Tuttavia, si nota dal modello che gli individui con atteggiamenti universalistici hanno una propensione molto elevata a votare anche la Sinistra radicale e socialdemocratici, seppur in misura minore rispetto al votare i Verdi.

Inoltre, come volevasi dimostrare, la dimensione Stato-mercato influisce maggiormente nella scelta dei partiti tradizionali: posizioni a favore del mercato tendono a far propendere il voto per i Liberali e per i Conservatori cristiani, ovvero per i gruppi di partito di destra moderata, mentre posizioni più favorevoli all'intervento dello Stato in termini di redistribuzione e di assistenza sociale sono associate a una propensione per i Socialdemocratici e per la Sinistra radicale.

Anche alcune variabili di controllo sono, per questa categoria di Paesi, fortemente indicative. Per esempio, l'età risulta il fattore chiave che spinge gli individui a scegliere i partiti tradizionali: più si è avanti con l'età e maggiore è la propensione a votare i partiti storici. Questo fenomeno potrebbe essere spiegato dal fatto che le nuove generazioni non hanno vissuto le divisioni sociali sulle quali sono nati questi partiti e quindi non ne comprendono i valori e le battaglie.

Anche la religione è ancora una significativa determinante di voto, sia in questo gruppo di Paesi che negli altri, dimostrando che il *cleavage* religioso non è ancora del tutto

metodologici per la raccolta dati. Per maggiori informazioni: <https://www.europeansocialsurvey.org/about/faq.html>.

⁹Associazione meno forte per altri gruppi di Paesi, per esempio per i Paesi orientati all'uguaglianza.

scomparso.

Per tutti i gruppi di Paesi, l'aumentare del reddito è associato a un aumento della probabilità di votare i partiti tradizionali di destra e, nei Paesi dell'Europa continentale e settentrionale, anche a un aumento della probabilità di votare i socialdemocratici, seppur con intensità minore.

Per la classe c'è da fare un discorso un po' diverso: nei Paesi orientati allo status, i voti dei più istruiti sono fortemente indirizzati ai Verdi, mentre, più in generale, negli altri gruppi di Paesi, questi soggetti non vedono di buon occhio la Sinistra, la Destra radicale e neanche i socialdemocratici.

Un'altra tendenza generale che emerge, è la crescente predisposizione dei professionisti socio-culturali a votare i Verdi.

Infine, il modello restituisce un'associazione che non dovrebbe sorprendere, ovvero che il voto dei sindacalisti è fortemente sbilanciato a sinistra, in tutte le sue sfaccettature, Verdi inclusi.

Tornando però alle due variabili indipendenti, sempre in [Häusermann and Kriesi, 2015], viene dimostrato che la dimensione universalismo-particolarismo è la dimensione che più influenza le scelte elettorali nei Paesi orientati alla cattura, allo status e alla competitività; solamente per i Paesi orientati all'uguaglianza prevale la dimensione Stato-mercato nel plasmare le preferenze elettorali e, per questo motivo, si può affermare che, a livello europeo, i temi riguardanti la dimensione che contrappone l'universalismo al particolarismo, siano i più importanti nel plasmare le preferenze politiche.

Quanto appena affermato è riscontrabile nei risultati elettorali dei partiti che hanno basato i rispettivi programmi e le rispettive campagne elettorali su questi temi.

Infatti, sia i Verdi che la Destra radicale hanno assunto un ruolo da protagonisti nel contesto politico europeo, toccando percentuali di oltre il 20% dei consensi in molti Paesi europei, entrando così a gamba tesa nelle istituzioni e nei governi delle principali democrazie europee [Häusermann and Kriesi, 2015] [Hooghe and Marks, 2018].

Nel capitolo 5, verrà effettuata un'analisi volta a individuare i fattori che influenzano il voto degli europei, utilizzando però dati più recenti, relativi all' ESS 2018, e un diverso modello statistico, il quale verrà spiegato nel capitolo 4. Esso sposterà l'attenzione su un'altra problematica che sorge quando si studiano fenomeni legati a processi decisionali dove i soggetti analizzati sono eterogenei e dispongono di set di scelte differenti.

3. Inferenza bayesiana

L'obiettivo di questo capitolo è quello di cercare di fornire una panoramica su una logica inferenziale che differisce dall'approccio classico e che prende il nome di *inferenza bayesiana*. Prima di entrare nel cuore di questo modo di intendere l'inferenza, è utile approfondire il teorema che ne è alla base, il *Teorema di Bayes*, e per farlo sarà necessario ricordare alcuni teoremi della probabilità.

Dopo aver spiegato questo importante teorema, verrà fatto un breve inciso per ricordare le principali caratteristiche dell'inferenza classica e poi si inizierà la trattazione dell'inferenza bayesiana. Si vedranno le caratteristiche principali di questo approccio, si evidenzieranno le differenze con l'approccio classico e infine verranno approfonditi diversi metodi di stima bayesiani.

3.1 Teorema di Bayes

Prima di introdurre il teorema di Bayes, è necessario ricordare il concetto di probabilità condizionata, la legge delle probabilità composte e il teorema delle probabilità totali.

Si definisce *probabilità condizionata* quella probabilità per la quale, dati due eventi A e B , si verifica l'evento A una volta verificatosi l'evento B . Nella fattispecie, si verifica quando l'evento B *condiziona* il verificarsi dell'evento A ; in termini matematici si rappresenta con $P(A|B)$.

Fissato quindi un evento $B \in \mathcal{A}$, contenuto in uno spazio di probabilità (Ω, \mathcal{A}, P) , la relazione che esprime la probabilità condizionata sarà:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3.1)$$

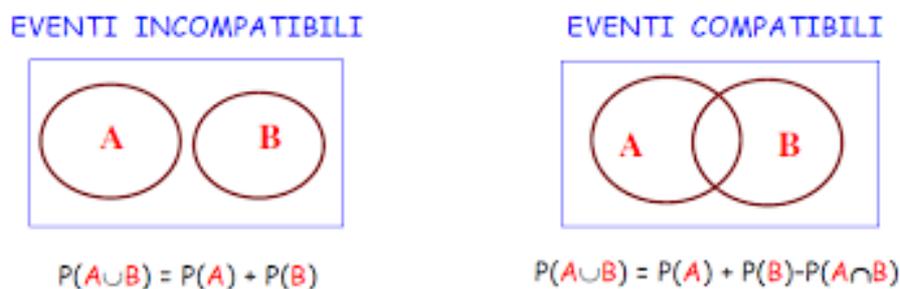


Figura 3.1: Diagramma di Venn. Eventi incompatibili e compatibili.

Da questa equazione è possibile derivare, sempre considerando che gli eventi non abbiano probabilità nulla, la *legge delle probabilità composte*, ovvero:

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \quad (3.2)$$

Questa relazione è valida nel caso in cui A e B siano due eventi dipendenti. Nel caso in cui i due eventi siano *stocasticamente indipendenti*, si avrà che $P(A|B) = P(A)$, cioè l'evento B non influenza la probabilità che si verifichi l'evento A; la relazione 3.2 si riduce in:

$$P(A \cap B) = P(A) \cdot P(B) \quad (3.3)$$

Passando al teorema delle probabilità totali, è necessario introdurre in prima istanza il concetto di eventi compatibili e incompatibili, per comprenderne la relazione che vi è alla base: *“Dati due eventi A e B, si dicono compatibili se il verificarsi dell'uno non esclude il verificarsi dell'altro e quando è possibile che si verifichino contemporaneamente. I due eventi si dicono invece incompatibili se il verificarsi dell'uno esclude il verificarsi dell'altro, ovvero quando non vi è la possibilità che si verifichino contemporaneamente”*.¹

In questo caso, la rappresentazione grafica contenuta nella figura 3.1² torna molto utile per rendere di più semplice comprensione questi due concetti. Come si può notare, nel primo caso i due eventi non si intersecano, sono quindi detti incompatibili, mentre nel secondo caso vi è un'intersezione tra i due insiemi, saranno dunque eventi compatibili.

¹<https://www.youmath.it/lezioni/probabilita/probabilita-discreta/1200-eventi-compatibili-e-incompatibili.html>.

²Diagramma di Venn.

Ora è possibile introdurre il teorema delle probabilità totali e declinarlo a seconda dei casi. Tale teorema permette di calcolare l'unione di due eventi³, ovvero assegnare una probabilità al verificarsi di uno o dell'altro.

Se i due eventi sono *compatibili*, cioè se $A \cap B \neq \emptyset$ e $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, allora si sottrae la probabilità dell'intersezione dei due eventi perché altrimenti verrebbe conteggiata due volte.

Nel caso di *eventi incompatibili*, si pensi a un evento e al suo complementare, avremo $A \cap A^c = \emptyset$ e di conseguenza $P(A \cup A^c) = P(A) + P(A^c)$, la probabilità dell'unione dei due eventi sarà semplicemente pari alla somma delle probabilità dei singoli eventi. Sia B un evento che può verificarsi solamente se associato ad uno dei due eventi precedenti, varrà allora la nota relazione che prende il nome di *teorema delle probabilità totali*, ovvero:

$$P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c) \quad (3.4)$$

Teorema di Bayes. Nel caso di un esperimento caratterizzato da eventi compatibili e dipendenti, dal teorema delle probabilità totali e dalla legge delle probabilità composte si deriva il *Teorema di Bayes*:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (3.5)$$

Gli eventi A e B devono essere probabili, vale a dire $P(A), P(B) > 0$. Nel caso in cui la probabilità che si verifichi l'evento B sia strettamente compresa tra zero e uno, la 3.5 potrà essere riscritta come:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)} \quad (3.6)$$

La 3.5 e la 3.6 rappresentano i casi più elementari poiché costruite per esperimenti caratterizzati da due eventi. Tuttavia, la versione della formula più comune nelle appli-

³Si può estendere anche a più di due eventi; per semplicità si è voluto considerare il caso più elementare.

cazioni reali prevede che vi siano “n” eventi incompatibili⁴ A_1, \dots, A_n , partizione finita o numerabile di Ω , in relazione con un evento B.

In tal caso:

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i) \quad (3.7)$$

da cui il teorema di Bayes diventa:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)} \quad (3.8)$$

Questa semplice ma fondamentale relazione, è alla base di un’intera branca della statistica, chiamata appunto *statistica bayesiana*, e presenta un’applicabilità straordinaria in tutti quei contesti reali dove vi sia necessità di aggiornare la probabilità di un dato evento alla luce di nuove informazioni.

Il termine $P(A_i|B)$ indica infatti la *probabilità a posteriori* di A_i , ovvero la probabilità che l’evento A_i (evento condizionato) sia vero, dato il verificarsi dell’evento B (evento condizionante). Per contestualizzare meglio tale teorema, potrebbe essere utile associare ad A_i le varie ipotesi o cause di un esperimento, mentre a B associarne l’evidenza o l’effetto.

Se con la probabilità condizionata si calcola la probabilità dell’evidenza date le ipotesi, con il teorema di Bayes si effettua l’operazione inversa, ovvero si calcola la probabilità che l’ipotesi sia vera data una certa evidenza.

$P(\text{Evidenza} | \text{Ipotesi}) \rightarrow \text{probabilità condizionata.}$

$P(\text{Ipotesi} | \text{Evidenza}) \rightarrow \text{Teorema di Bayes.}$

Se un effetto può essere provocato da diverse cause, l’equazione 3.8 ci permette di calcolarne le relative probabilità.

Analizzando il secondo termine di tale equazione, il denominatore indica la *probabilità a priori* di osservare una certa evidenza senza fare assunzioni sulle ipotesi. Al numeratore, il termine $P(A_i)$ indica la *probabilità a priori* che si verifichi una determinata ipotesi senza l’influenza dell’evidenza B; è un dato già esistente che dipende dalle informazioni pregresse di colui che effettua l’esperimento.

⁴Ovvero $A_i \cap A_j = 0$, per ogni $i \neq j$.

Il termine $P(B|A_i)$ indica invece la *verosimiglianza* delle ipotesi A_i , ovvero la probabilità di dare l'evidenza B se l'ipotesi A è vera. Questo termine rappresenta bene il concetto di probabilità soggettiva presentato nel paragrafo A.1.2.3, in quanto esprime quanto scommetterebbe il valutatore sull'evidenza conoscendo le ipotesi o, in altri termini, quanto scommetterebbe sugli effetti di un certo esperimento conoscendone le cause.

Per concretizzare tale teorema, procediamo con un esempio. Questo lavoro è stato elaborato durante la crisi sanitaria causata dal virus COVID-19 e, per questo motivo, l'esempio prenderà a riferimento l'accuratezza dei test sierologici, comunemente chiamati "tamponi", nel diagnosticare l'infezione⁵.

Si ipotizzi che in Italia il tampone risulti positivo nel 98% dei casi se il virus è presente e nel 3% dei casi se assente. Il virus risulta inoltre essere presente in 33 persone su 10 000⁶. L'obiettivo dello studio è quello di capire la probabilità che un individuo sia effettivamente malato se risulta positivo a un test.

La lettera "A" sarà associata al verificarsi dell'evento "l'individuo è malato" (ipotesi), mentre la lettera "B" al verificarsi dell'evento "il test è risultato positivo" (evidenza). Quindi: $P(A) = 0.0033$, $P(A^c) = 0.9967$, $P(B|A) = 0.98$, $P(B|A^c) = 0.03$.

Calcoliamo dunque, attraverso la formula di Bayes, la probabilità che l'individuo abbia contratto il virus risultando positivo al tampone:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)} = 0.0976 \quad (3.9)$$

Come si può notare, tale probabilità è molto bassa; nonostante il tampone restituisca la positività al virus, è estremamente improbabile che l'individuo oggetto di analisi sia effettivamente malato. Ciò denota una scarsa precisione del tampone, in quanto restituisce un'elevatissima percentuale di falsi positivi [Caravenna and Dai Pra, 2013].

Questo teorema permette, dunque, per quanto ci concerne, di fare inferenza statistica su quantità non note in base all'evidenza disponibile e a un'ipotesi iniziale; è proprio questo l'elemento che ricollega il concetto di soggettività, analizzato nella paragrafo A.1.2.3, alle analisi bayesiane [Iannazzo, 2007].

⁵I valori riferiti agli *output* dei tamponi sono frutto d'immaginazione, non hanno alcuna valenza scientifica.

⁶Valore ottenuto rapportando il numero di casi positivi [John Hopkins] al 30/04 in Italia sul numero totale di abitanti [ISTAT].

3.2 Approccio inferenziale bayesiano

L'*inferenza* è la generalizzazione dei risultati ottenuti attraverso la rilevazione per campione a dei collettivi più ampi detti "popolazione". Si tratta di uno strumento statistico utilizzato laddove non sia possibile analizzare l'intera popolazione oggetto di studio, in quanto si tratta di un'operazione generalmente molto costosa in termini economici e di tempo [Ballone and Colagrande, 1994].

La tecnica utilizzata è quella di prendere a riferimento un *campione aleatorio*, ovvero la realizzazione numerica di un sottoinsieme di "n" unità statistiche rappresentative della popolazione e studiarlo con l'obiettivo di stimare i parametri non noti della popolazione. In questo modo il campione è costruito in modo tale che sia identicamente e indipendentemente distribuito rispetto alla popolazione, permettendo appunto l'inferenza sui parametri non noti.

Si ricordi che la popolazione di riferimento è una *variabile aleatoria* X definita sulla sua *funzione di ripartizione* $F(x, \theta)$ e si indica con $X \sim F(x, \theta)$, dove θ è un vettore di parametri fissi e incogniti. Da questa popolazione viene estratto casualmente un *sottoinsieme di unità statistiche* (X_1, X_2, \dots, X_n) , anch'esse variabili aleatorie e aventi la stessa distribuzione di X , la cui realizzazione numerica (x_1, x_2, \dots, x_n) costituisce il *campione aleatorio*.

Approccio classico all'inferenza. L'impostazione oggettivista, frequentista, nonché classica, basa il processo inferenziale sulla ripetibilità degli eventi analizzati, in modo che il campione in analisi rappresenti tutti possibili campioni ottenibili ripetendo l'esperimento infinite volte e sotto le stesse condizioni. Per estendere le informazioni ottenute dal campione così costruito alla popolazione, per minimizzarne le distorsioni e aumentarne la precisione, ci si avvale di tre metodologie. Esse si elencheranno brevemente per fornire una panoramica generale ma, essendo riferite all'inferenza classica, non saranno approfondite [Azzalini, 2004].

Il primo strumento, chiamato *stimatore*, è una funzione di un campione di variabili aleatorie volto a stimare il vettore di parametri non noti della popolazione⁷. Esso è dunque una variabile aleatoria T funzione del campione: $T = (X_1, X_2, \dots, X_n)$. La

⁷Questa procedura prende il nome di *teoria della stima*.

distribuzione campionaria dello stimatore è detta *legge campionaria* e dipende dai vettori di parametri non noti, mentre lo stimatore T non dipende mai da θ . La *stima* rappresenta invece una delle possibili realizzazioni dello stimatore e restituisce un valore numerico [Epifani, 2009]. La seconda tecnica è la *verifica delle ipotesi statistiche*, o *test d'ipotesi*. Essa viene utilizzata per verificare, confrontando due tesi complementari riguardanti θ , quale sia la più plausibile sulla base delle informazioni desunte dal campione [Ballone and Colagrande, 1994].

Infine, molto utilizzata è la *teoria degli intervalli di confidenza*. Alla stima puntuale, in quanto aleatoria, si aggiunge un margine d'errore, creando così un intervallo di valori reali in cui riporre una prefissata fiducia per θ [Azzalini, 2004].

Approccio inferenziale bayesiano. L'inferenza classica si basa su un concetto oggettivo di probabilità strettamente legato all'evento, presuppone la conoscenza del modello rappresentativo del fenomeno⁸ e ha come obiettivo la stima dei parametri non noti della popolazione, considerati come delle costanti, avvalendosi dei dati campionari [James et al., 2013].

L'inferenza bayesiana poggia le sue basi su una logica differente. In *primis*, è legata a un concetto di *probabilità soggettiva*⁹, assume centralità la figura di colui che effettua l'esperimento, ossia le sue conoscenze pregresse, in quanto la probabilità è intesa come il *grado di fiducia* che egli associa al verificarsi di un dato evento.

Inoltre, considera i parametri *variabili aleatorie*, contrariamente all'approccio classico che li tratta come fissi. Essendo considerati variabili aleatorie e non più costanti, anch'essi saranno governati da una loro legge di distribuzione ed è qui che entra in gioco l'esperienza del ricercatore: le sue conoscenze pregresse del fenomeno si risolvono nell'ipotizzare una *distribuzione di probabilità a priori* per i parametri, ancor prima di analizzare i dati del campione.

Gli *output* ottenuti dallo studio dei dati campionari x serviranno a fornire nuove informazioni sulle componenti non note, ovvero serviranno ad aggiornare la situazione di partenza delineata da colui che effettua lo studio. In altri termini, l'analisi baye-

⁸Sia nel caso parametrico che in quello non parametrico, anche la forma analitica del modello è incognita e i dati campionari verranno utilizzati per stimare sia i parametri che la forma del modello.

⁹Si veda il paragrafo A.1.2.3.

siana cerca di aggiornare le conoscenze pregresse riguardanti parametri non noti della popolazione θ utilizzando i dati campionari.

L'obiettivo è quello di determinare la *distribuzione di probabilità a posteriori* dei parametri incogniti: essa esprime quanto il ricercatore scommetterebbe sui possibili valori dei parametri dopo aver esaminato i dati campionari rappresentativi del fenomeno analizzato. La formula di Bayes, per come è strutturata, esprime in maniera ottimale il *processo di apprendimento* dell'esperienza e risulta quindi essere la più indicata a tale scopo. Si assuma una variabile casuale $X \sim f(x, \theta)$ ¹⁰. Considerando le osservazioni $x = (x_1, \dots, x_n)$ e i parametri $\theta = (\theta_1, \dots, \theta_n)$, l'equazione 3.5 si può riscrivere come:

$$P_{\Theta|X}(\theta|x) = \frac{P_{X|\Theta}(x|\theta) \cdot P_{\Theta}(\theta)}{P_X(x)} \quad (3.10)$$

La *verosimiglianza marginale* $P(x)$ può essere ottenuta integrando il numeratore rispetto a θ , ottenendo così:

$$P_X(x) = \int P_{X|\Theta}(x|\theta) \cdot P_{\Theta}(\theta) d(\theta) \quad (3.11)$$

In questo modo la 3.10 diventa:

$$P_{\Theta|X}(\theta|x) = \frac{P_{X|\Theta}(x|\theta) \cdot P_{\Theta}(\theta)}{\int P_{X|\Theta}(x|\theta) \cdot P_{\Theta}(\theta) d(\theta)} \quad (3.12)$$

$P_{\Theta|X}(\theta|x)$ è la distribuzione a posteriori ed esprime la probabilità che i parametri assumano determinati valori, dopo aver fatto su di essi delle assunzioni e dopo aver effettuato l'esperimento. Rispecchia quindi la scommessa che il valutatore farebbe sui possibili valori dei parametri qualora conoscesse gli esiti dello studio campionario. In questo modo, lo studioso può aggiornare le proprie conoscenze pregresse utilizzando le informazioni estratte dallo studio del campione in esame. Questo permette di aggiornare, con la ricerca in atto, il patrimonio informativo antecedente, arricchendolo e valorizzandolo. Questa tecnica di aggiornamento è alla base di alcuni strumenti¹¹ utilizzati nel

¹⁰Si faccia attenzione all'uso della simbologia utilizzata: in questo caso si è utilizzata la virgola come elemento separatore delle due entità dell'argomento, cioè $f(x, \theta)$, in quanto entrambe variabili; se l'elemento separatore fosse stato il punto e virgola, cioè $f(x; \theta)$, la prima entità sarebbe stata da considerare variabile mentre la seconda costante.

¹¹Ad esempio le reti bayesiane, per maggiori approfondimenti [Grossman and Domingos, 2004],

machine learning: proprio per la sua struttura, è infatti possibile indurre un processo di autoapprendimento tramite l'implementazione di appositi algoritmi.

$P_{X|\Theta}(x|\theta)$, chiamata *funzione di verosimiglianza*¹² o *probabilità condizionata del campione*, rappresenta la *distribuzione di probabilità* che lo sperimentatore assegnerebbe ai dati campionari dato il parametro o i parametri. Da qui si può già notare il *livello di soggettività*: questa distribuzione non è altro che l'esito di una *scommessa* che un *soggetto razionale* farebbe sui dati campionari conoscendo i parametri d'interesse; dato che un certo θ è vero, si vedranno delle x , dei dati, aventi una determinata forma.

$P_{\Theta}(\theta)$, chiamata *probabilità a priori*, è la *distribuzione di probabilità a priori dei parametri*: deriva dall'esito della scommessa sui parametri che il valutatore formula ancor prima di aver analizzato il campione e aver estratto da esso informazioni utili sul fenomeno in questione. Il parametro non è più fisso ma è anch'esso governato da una legge di distribuzione della probabilità, risultando così una variabile aleatoria.

Anche in questo caso c'è un forte *tratto soggettivo* associato alle valutazioni che l'esperto formula sulla distribuzione di probabilità dei parametri; questo tratto è ancor più forte se si pensa che esse sono formulate senza nessuna informazione sul campione ma solo in base all'esperienza e alle conoscenze pregresse del fenomeno.

Il denominatore, $P_X(x)$, chiamato *distribuzione predittiva a priori di x* , definisce la *distribuzione marginale* di $x = (x_1, \dots, x_n)$. Questa distribuzione dipende esclusivamente dal campione e funge da *fattore di normalizzazione*. Per la stima dei parametri è semplicemente una costante atta a garantire che $P(\theta|x)$ sia ancora una distribuzione: grazie a $P(x)$ l'integrale della distribuzione di probabilità a posteriori dà come risultato 1. Nell'inferenza bayesiana, quando si usano metodi di simulazione MCMC, questo termine può essere ignorato e l'equazione 3.10 diventa:

$$P_{\Theta|X}(\theta|x) \propto P_{X|\Theta}(x|\theta) \cdot P_{\Theta}(\theta) \quad (3.13)$$

Basta dunque una formula in cui la *posterior* sia proporzionale al prodotto tra la *prior* e la verosimiglianza.

Infine, l'equazione 3.10 è una generalizzazione della regola di Bayes, in quanto quest'ultima ha diverse varianti a seconda che i dati trattati siano discreti o continui così come

[De Grassi et al., 2009], [Friedman et al., 1997].

¹²Può essere riscritta come $\prod_{n=1}^n P(x_n|\theta)$.

le quantità da stimare [Rossi et al., 2012] [Lesaffre and Lawson, 2012].

Problema del test d'ipotesi. Nel problema del test d'ipotesi, la quantità sconosciuta Θ è discreta e quindi l'*output* del teorema di Bayes sarà una PMF, cioè una p_{Θ} , sia con dati discreti che continui.

Nel caso in cui sia i dati siano discreti l'equazione sarà:

$$p_{\Theta|X}(\theta|x) = \frac{p_{X|\Theta}(x|\theta) \cdot p_{\Theta}(\theta)}{p_X(x)} \quad (3.14)$$

Nel caso di dati continui prenderà, invece, questa forma:

$$p_{\Theta|X}(\theta|x) = \frac{f_{X|\Theta}(x|\theta) \cdot p_{\Theta}(\theta)}{f_X(x)} \quad (3.15)$$

Si può notare che nel caso completamente discreto tutte le quantità sono descritte dalle PMF mentre, nel caso con i dati continui, le x sono descritte dalle PDF [Tsitsiklis, 2010] [Bertsekas and Tsitsiklis, 2008].

Problemi di stima. Nei problemi di stima, la quantità da stimare è continua: la *posterior* sarà descritta da una PDF, da una f_{Θ} , sia con dati discreti che con dati continui.

Se i dati sono continui si avrà:

$$f_{\Theta|X}(\theta|x) = \frac{f_{X|\Theta}(x|\theta) \cdot f_{\Theta}(\theta)}{f_X(x)} \quad (3.16)$$

Mentre con dati discreti:

$$f_{\Theta|X}(\theta|x) = \frac{p_{X|\Theta}(x|\theta) \cdot f_{\Theta}(\theta)}{p_X(x)} \quad (3.17)$$

Un esempio di problemi di stima con dati continui può essere il seguente. Si immagina di lanciare un pallone da football e di volerne modellare la traiettoria. Essendo la traiettoria del lancio soggetta alla gravità, essa sarà riconducibile a una parabola. L'equazione generale di una parabola è: $z_t = \Theta_0 + t\Theta_1 + t^2\Theta_2$, dove z_t è la posizione del pallone al momento "t". Non si conosce esattamente la forma della parabola, in

quanto i parametri Θ della parabola sono incogniti. Si cerca allora di tracciare, con appositi strumenti, la posizione del pallone in movimento in vari momenti. Le misurazioni sono però disturbate, cioè sono imprecise; avremo dunque $x_t = z_t + \epsilon_t$, ovvero la vera posizione del pallone sarà pari alla relativa misurazione più un certo errore sistematico. Quindi, sulla base dei dati ottenuti dalle misurazioni, si vuole scoprire l'esatta equazione della parabola relativa al lancio del pallone attraverso la miglior stima dei Θ .

Essere a conoscenza delle distribuzione di probabilità dei parametri coincide con l'essere a conoscenza della *prior*, quindi bisognerà avere a disposizione le distribuzioni dei Θ , cioè le $f_{\Theta}(\theta)$, per poter continuare con l'analisi.

Si passa poi alla funzione di verosimiglianza, alla distribuzione condizionata delle x dati i θ . Se si conosce la distribuzione condizionata di z_t dati i Θ e si utilizza l'equazione di x_t , si può ricavare la distribuzione di X quando θ assume un determinato valore.

A questo punto si hanno gli elementi necessari a stimare, tramite gli algoritmi MCMC presentati nella sezione 3.3.3, la distribuzione a posteriori dei parametri:

$$f_{\Theta_0, \Theta_1, \Theta_2 | X_1, \dots, X_n}(\theta_0, \theta_1, \theta_2 | x_1, \dots, x_n) \quad (3.18)$$

Si noti che X e Θ non sono singole variabili aleatorie, ma variabili aleatorie multidimensionali. Nel nostro caso la quantità da stimare corrisponderà a una tripletta, a un vettore $\Theta_0, \Theta_1, \Theta_2$. Anche la X , in questo caso, corrisponde all'intera sequenza di misurazioni x_1, \dots, x_n effettuate.

Per fare un esempio di problemi di stima con dati discreti, si pensi di lanciare una moneta infinite volte. La moneta presenta una distorsione per l'uscita di "testa" catturata dal parametro θ . Si osservano X teste in "n" lanci; cosa si può dire sul vero valore di θ ? Nell'inferenza classica si userà uno stimatore non distorto per la stima del parametro, ovvero:

$$\hat{\Theta}_n = \frac{X}{n} \quad (3.19)$$

Questo particolare stimatore consiste nel rapportare il numero di teste ottenute in n prove con il numero totale di prove. A questo punto si dimostrerebbero alcune buone proprietà dello stimatore, per esempio la convergenza in probabilità al vero valore del parametro, e il problema sarebbe finito.

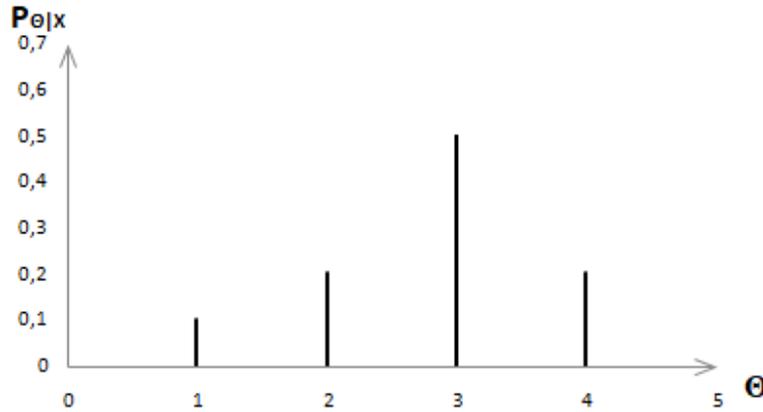


Figura 3.2: Esempio distribuzione a posteriori nel problema con Θ discreti.

Nell’approccio bayesiano, la distorsione della moneta non viene considerata una costante sconosciuta da ricavare, ma una variabile aleatoria con una propria distribuzione. Questa distribuzione è ciò che abbiamo chiamato distribuzione a priori. Per esempio, si potrebbe assumere che i vari θ siano equiprobabili, che tutti i valori per la distorsione della moneta abbiano la stessa probabilità, così da assumere una *prior* uniforme.

Si potrebbe invece pensare che la moneta sia non distorta, ovvero che sia fabbricata in modo tale che, a seguito di un lancio, vi sia il 50% di ottenere testa e il 50% di ottenere croce. La moneta potrebbe però non essere perfettamente bilanciata, cioè potrebbe avere qualche piccolo difetto di fabbricazione; la distribuzione di una serie di lanci potrebbe essere centrata su $\frac{1}{2}$ in quanto, in linea di massima, non è distorta, ma potrebbe comunque presentare un minimo di dispersione causata dagli errori di fabbrica.

Si assume quindi la *prior* ritenuta più coerente con l’esperimento e si calcola poi la probabilità dei diversi valori di Θ alla luce dei dati osservati [Bertsekas and Tsitsiklis, 2008] [Tsitsiklis, 2010].

Stima del massimo a posteriori. Si pensi al problema del test d’ipotesi: la quantità da stimare è discreta e il *plot* relativo, a prescindere che i dati siano continui o discreti, sarà descritto da un grafico simile a quello illustrato nella figura 3.2.

Sulle ascisse vi sono i Θ , i possibili valori per la quantità sconosciuta, i quali sono

trattati come V.A. discrete. Sulle ordinate troviamo le PMF a posteriori relative ai diversi valori che la quantità da stimare potrebbe assumere, i quali restituiscono la probabilità che quel determinato valore sia quello corretto, una volta osservati i dati.

Osservando un grafico di questo tipo, il modo migliore di interpretarlo sarebbe dire che non si è a conoscenza del vero valore del parametro poiché i dati sono troppo limitati ma, in questo esempio, potrebbe essere il primo con probabilità del 10%, il secondo con probabilità del 20%, il terzo con probabilità del 50% e l'ultimo con probabilità del 20%.

Di solito si è però interessati a identificare un solo valore. Si cercherà allora la soluzione più probabile, ovvero di minimizzare la probabilità di prendere una decisione errata. Nell'esempio riportato nella figura 3.2, il valore più probabile è quello relativo a Θ pari a 3 e quindi si sceglierà questo valore. Così facendo si ha una probabilità di aver preso una decisione errata del 50%, mentre negli altri casi si sarebbe presa una decisione errata con probabilità pari all'80% per il secondo e l'ultimo parametro, mentre del 90% per il primo.

Il metodo che seleziona dalla posteriori il punto che presenta la probabilità più alta e che minimizza la probabilità di inferire in modo errato, è chiamato *stima del massimo a posteriori*, abbreviato M.A.P., tradotto in formula: $p_{\Theta|X}(\theta^*|x) = \max_{\theta} p_{\Theta|X}(\theta|x)$

Se per parametri che assumono valori discreti questa è una tecnica ragionevole, per i problemi di stima caratterizzati da parametri che assumono valori nel continuo non è la soluzione migliore. Si troverebbe il massimo assoluto della densità a posteriori, il quale però non è il punto più probabile come nel caso discreto, ma il punto in cui la densità, ossia la probabilità di cadere in un piccolo intervallo, è maggiore. Le probabilità relative ai singoli valori che il parametro può assumere sono pari a zero, quindi utilizzare il M.A.P. risulterebbe piuttosto inopportuno [Bertsekas and Tsitsiklis, 2008] [Tsitsiklis, 2010].

Stima LMS. Come appena accennato, per i problemi di stima il metodo M.A.P. non è molto ragionevole. Il grafico relativo a questo tipo di problema, sia che i dati siano continui che discreti, avrebbe una forma simile alla funzione rappresentata nella figura 3.3.

Si noti che il metodo M.A.P. ci porterebbe a scegliere il parametro in un intervallo centrato sul valore relativo al massimo assoluto, al secondo picco. Essendo una densità,

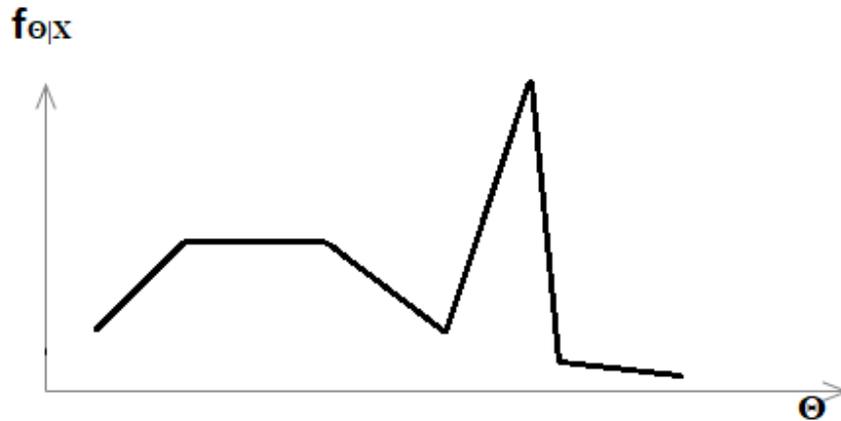


Figura 3.3: Esempio di una distribuzione a posteriori relativa a problemi di stima.

c'è però un'elevata densità di probabilità anche nell'intervallo relativo alla prima parte della funzione: vi è quindi un'elevata probabilità che il vero valore del parametro non si trovi sotto l'area relativa al massimo assoluto.

Con la premessa che anche in questo caso sarebbe preferibile ragionare sulla distribuzione a posteriori nella sua totalità, se si dovesse scegliere un valore specifico per il parametro, la strada più ragionevole sarebbe quella di calcolare il valore atteso della *posterior*, ovvero:

$$E[\theta|X = x] = \int \theta f_{\Theta|X}(\theta|x) d\theta \quad (3.20)$$

Il valore trovato sarà una sorta di centro di gravità per la distribuzione, quindi sarà la scelta più equilibrata per il parametro d'interesse. In effetti, effettuando una stima LMS, si nota che il valore atteso condizionato della distribuzione a posteriori è la miglior stima possibile per mantenere l'errore quadratico medio il più piccolo possibile.

Un esempio di stima LMS potrebbe essere il seguente: si pensi di dover affrontare un problema di stima senza avere nessuna informazione in più oltre alla distribuzione a priori. Si è convinti che la *prior* sia uniformemente distribuita in un intervallo compreso tra i valori 10 e 22 con probabilità pari a $\frac{1}{12}$, come osservabile nella figura 3.4.

Essendo interessati all'ottimizzazione dell'errore quadratico medio, si pensi che g sia un determinato stimatore e che $g(x)$ sia la sua realizzazione, la stima puntale. Se θ è il vero valore del parametro, allora $\theta - g(x)$ sarà l'errore di stima. Se l'obiettivo è la minimizzazione dell'errore quadratico medio, ovvero $\min E[(\theta - g(x))^2]$, usando la

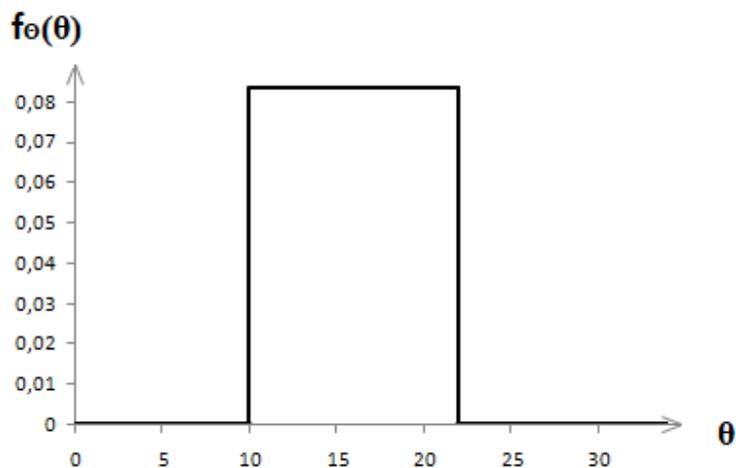


Figura 3.4: Densità a priori uniforme.

linearità delle aspettative si otterrà:

$$E[(\theta - g(x))^2] = E[\theta^2] - 2g(x)E[\theta] + g(x)^2 \quad (3.21)$$

Prendendo la derivata del secondo termine rispetto a $g(x)$, otterremo la quantità che minimizza l'errore quadratico medio e si nota che è:

$$E[\theta] = g(x) \quad (3.22)$$

Da ciò si evince che la stima che minimizza al meglio l'errore quadratico medio è il valore atteso di θ ; nell'esempio riportato nel grafico sarebbe $\theta^* = 16$. Lo stimatore migliore è quindi:

$$\hat{\theta} = E[\theta] \quad (3.23)$$

Se si è interessati alla dimensione media dell'errore commesso, sostituendo la stima generica $g(x)$ con lo stimatore $E[\theta]$ nella formula dell'errore quadratico medio, si scopre che:

$$E[(\theta - E[\theta])^2] = \text{var}(\theta) \quad (3.24)$$

Si noti che la dimensione media dell'errore non è nient'altro che la varianza della distribuzione.

Ora si vedrà come cambia l'analisi quando, oltre alla distribuzione a priori, si hanno a di-

sposizione anche dei dati relativi all'esperimento, ossia le X . Questo sposta l'analisi dalla distribuzione originale di θ , la distribuzione a priori, alla distribuzione condizionata di θ sui dati a disposizione, ovvero alla distribuzione a posteriori.

Si lavorerà dunque in un nuovo universo, un universo condizionato nel quale le probabilità non differiscono dalle probabilità che governano l'universo originale, se non per il fatto che vi sono nuove informazioni da tenere in considerazione.

Si avrà quindi uno scenario che presenta le due V.A. θ e X ; si osserva che $X = x$ e ci si sposta in un universo condizionato su quanto osservato. La cosa interessante, ma che a questo punto non dovrebbe stupire, è che la quantità ottimale per minimizzare l'errore quadratico medio è il valore atteso della distribuzione nel nuovo universo condizionato. Ripetendo la procedura fatta in precedenza, si nota infatti che:

$$\min E[(\theta - g(x))^2 | X = x] \quad (3.25)$$

Si deriva dunque rispetto a $g(x)$ e si ottiene:

$$g(x) = E[\theta | X = x] \quad (3.26)$$

La stima ottima è nuovamente il valore atteso della distribuzione, ma in un universo nel quale siamo a conoscenza dei risultati di un certo esperimento.

Quello che la formula 3.26 comunica è che non importa quale specifica X si sia verificata, per ogni X il valore atteso condizionato è la stima migliore.

Lo stimatore g che meglio ottimizza $E[(\Theta - g(x))^2]$ sarà quindi:

$$\hat{\theta} = E[\theta | X] \quad (3.27)$$

Se si osserva un valore specifico per la variabile aleatoria X , $X = x$, la miglior stima che minimizza l'errore quadratico medio sarà, per ogni x :

$$g(x) = E[\theta | X = x] \quad (3.28)$$

Inoltre, anche in questo caso, la dimensione dell'errore di questa stima è pari alla varianza della distribuzione a posteriori.

Riprendendo l'esempio precedente, si immagini che, oltre alla *prior*, si facciano delle osservazioni e si arrivi alla conclusione che, dato il valore di θ , X è una variabile aleatoria che oscilla tra $\theta - 1$ e $\theta + 1$; ovvero $X = \theta + u$ con u uniforme tra 0 e 1.

Si avrà dunque il vero valore del parametro, e assumere che la X sia uniformemente distribuita sull'intervallo $[\theta - 1; \theta + 1]$ è equivalente ad assumere che u sia anch'essa uniformemente distribuita su quell'intervallo. X sarà dunque una misurazione disturbata di θ rappresentata dal seguente grafico:

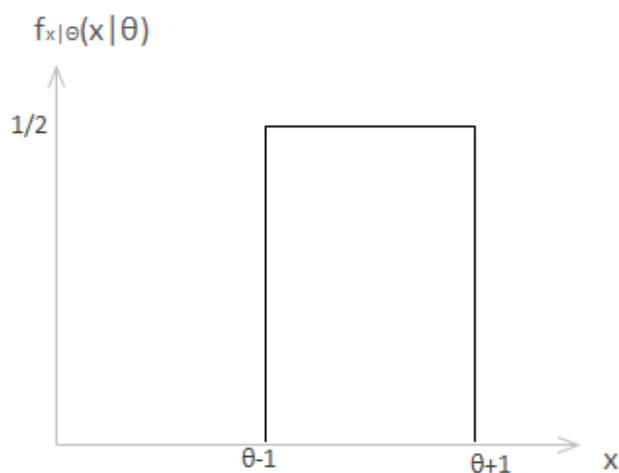


Figura 3.5: Densità condizionata.

Si hanno dunque tutti gli elementi per calcolare e plottare la densità congiunta, in quanto è pari alla densità a priori per la densità condizionata. Si avrà quindi che: $f_{X,\Theta}(x, \theta) = f_{\Theta}(\theta) \cdot f_{X|\Theta}(x|\theta) = \frac{1}{12} \cdot \frac{1}{2} = \frac{1}{24}$.

Si avrà una densità congiunta pari a $\frac{1}{24}$ sull'intervallo delle possibili x e dei possibili θ , mentre al di fuori di questi intervalli si avrà una densità pari a zero, come osservabile nella figura 3.6.

Una volta ottenuta la densità congiunta, l'insieme di tutte le possibili coppie (x, θ) , si procede cercando di costruire lo stimatore per θ in funzione delle x , ossia in funzione del fatto che lo stimatore ottimo è il valore atteso condizionato del parametro dato il valore di X osservato.

Condizionando quindi sulla x osservata, θ potrà essere compreso solamente tra i due estremi dell'intervallo relativo sulle ordinate¹³.

¹³Nel grafico, i due estremi relativi a x sono θ_1 e θ_2 .

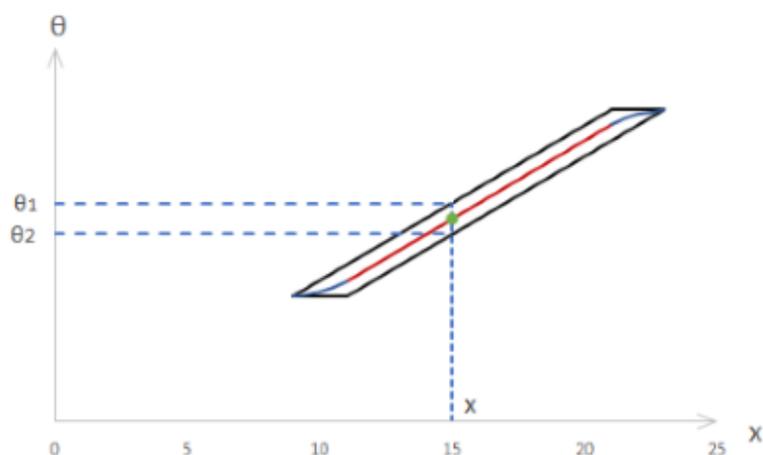


Figura 3.6: Densità congiunta e stimatore per il parametro d'interesse.

Riprendendo le relazioni citate nella trattazione delle variabili aleatorie multiple nell'appendice B.1, la distribuzione di θ date le x , sarà una sezione della distribuzione congiunta delle due V.A. applicata alla x sulla quale si sta condizionando. Essendo la distribuzione congiunta costante su quel determinato intervallo, anche la *posterior* di θ sarà costante in quell'intervallo e, di conseguenza, il valore atteso di θ per ogni x sarà il punto medio di ogni relativo intervallo. Unendo tutti i punti medi, tutti i valori attesi condizionati, si plotterà il nostro stimatore che, in questo caso, così come in molti altri, sarà uno stimatore non lineare¹⁴. L'*output* sarà quindi una stima $g(x)$ che non ha nessun motivo di essere una funzione lineare di x .

Una volta ricavato il valore atteso condizionato si procede con l'errore quadratico medio. Avendo osservato un particolare valore di X , si è nell'universo in cui $X = x$, allora si avrà che:

$$E[(\theta - E[\theta|X])^2|X = x] = \text{var}(\theta|X = x) \quad (3.29)$$

Questa varianza è dunque la varianza della *posterior* di θ date le osservazioni effettuate. Nell'esempio fatto, si è detto che il parametro è uniforme su un determinato intervallo di lunghezza pari a 2 e questa è la distribuzione a posteriori del parametro stesso. La varianza per una distribuzione uniforme è data dalla lunghezza dell'intervallo al

¹⁴Il valore atteso condizionato è uno stimatore lineare solo se tutte le variabili aleatorie sono normali.

quadrato diviso 12, quindi sarà pari a $\frac{1}{3}$.

Questa varianza caratterizzerà le x comprese tra 11 e 21, mentre agli estremi, per le x comprese tra 9 e 11 e tra 21 e 23, la varianza si comporterà un po' diversamente. Per le x pari a 9 e a 23, la varianza sarà nulla perché i rispettivi valori di θ sono esattamente 10 e 22. Ciò sta a significare che alcune osservazioni sono migliori di altre, vedere x pari a 9 o a 23 significa conoscere esattamente il vero valore del parametro.

Nell'intervallo delle x comprese tra 9 e 11, la varianza partirà da 0 fino ad arrivare a $\frac{1}{3}$, mentre partendo da 21 fino ad arrivare a 23 partirà da $\frac{1}{3}$ per arrivare a 0. Sicuramente un'interpretazione visiva può risultare di più facile comprensione:

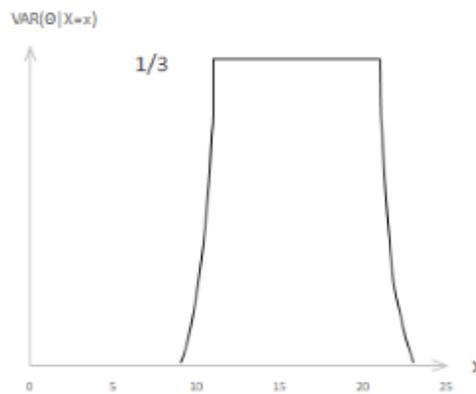


Figura 3.7: Distribuzione errori quadratici medi.

La varianza dell'errore dipende quindi da quale particolare osservazione capita di rilevare: alcune danno molte informazioni sul parametro d'interesse mentre altre potrebbero lasciare molti interrogativi.

Un ulteriore passo in avanti è capire se esiste una relazione sistematica, una covarianza, tra errore e stima. A tale scopo è utile esplicitare le principali proprietà delle stime LMS, ovvero:

1. $\hat{\theta} = E[\theta|X]$ è lo stimatore;
2. $\tilde{\theta} = \hat{\theta} - \theta$ è l'errore di stima;
3. $E[\tilde{\theta}] = 0$; in media gli errori si annullano tra loro;

$$\frac{15 \cdot 2^2}{12}$$

4. $E[\tilde{\theta}|X] = E[\hat{\theta} - \theta|X] = E[\hat{\theta}|X] - E[\theta|X] = \hat{\theta} - \hat{\theta} = 0$, grazie alla linearità delle aspettative che permette la scomposizione del valore atteso, si scopre che il valore atteso dell'errore di stima è pari a 0. $E[\theta|X]$ è pari a $\hat{\theta}$ per definizione¹⁶, mentre $E[\hat{\theta}|X] = \hat{\theta}$ poiché, essendo la stima effettuata sulla base dei dati, delle X, $\hat{\theta}$ è una funzione di X quindi, nota la X, anche $\hat{\theta}$ sarà noto.

Si noti anche che $E[\tilde{\theta}|X = x] = 0$, ovvero che non importa quale X venga osservata, in media l'errore di stima è pari a 0.

Gli stimatori che, in media, presentano un errore di stima nullo sono chiamati *stimatori non distorti*, poiché restituiscono stime in cui gli errori positivi e quelli negativi in media hanno lo stesso peso, cioè sono bilanciati;

5. $E[\tilde{\theta}, h(X)|X] = h(X) \cdot E[\tilde{\theta}|X] = 0$. Si noti che anche nel caso in cui l'errore di stima sia collegato a una certa funzione dei dati, il relativo valore atteso è pari a 0.

Essendo in un universo condizionato nel quale si conoscono i dati, ossia la X, si conoscerà anche la forma di $h(X)$: non essendo casuale, si può trattare come una costante e quindi scorporarlo dal valore atteso.

Prendendo il valore atteso su entrambi i termini nell'equazione appena vista e utilizzando la legge delle aspettative ripetute¹⁷, si avrà che $E[\tilde{\theta}, h(X)] = 0$.

La covarianza è pari al valore atteso del prodotto delle due variabili aleatorie meno il prodotto dei valori attesi delle variabili aleatorie prese singolarmente, quindi:

$$Cov(\tilde{\theta}, h(X)) = E[(\tilde{\theta}, h(X))] - E[\tilde{\theta}] \cdot E[h(X)] = 0 - 0 = 0 \quad (3.30)$$

Il risultato ottenuto è facilmente verificabile applicando all'equazione della covarianza le proprietà dei valori attesi condizionati citate nei punti 3 e 5.

Nel caso in cui, al posto di una funzione generica dei dati, ci sia lo stimatore $\hat{\theta}$, si avranno le stesse conclusioni della formula precedente, ovvero:

$$Cov(\tilde{\theta}, \hat{\theta}) = 0 \quad (3.31)$$

¹⁶Si guardi il punto uno.

¹⁷Questa legge dice che il valore atteso di un valore atteso condizionato restituisce il valore atteso nell'universo non condizionato.

La covarianza tra errore e stima è quindi pari a zero. Questa conclusione è di fondamentale importanza perché evidenzia che tra errore e stima non c'è nessun tipo di legame sistematico, non sono correlati. In altri termini, uno stimatore non dovrebbe produrre stime che siano sistematicamente maggiori o minori rispetto al vero valore del parametro, ma stime a volte maggiori e a volte minori.

Di conseguenza, conoscere $\hat{\theta}$ non restituisce informazioni sugli errori ed è lo stimatore ottimo in termini di minimizzazione dell'errore quadratico medio, in questi termini non vi è uno stimatore migliore. Il vero valore del parametro sarà pari a:

$$\theta = \hat{\theta} - \tilde{\theta} \quad (3.32)$$

Essendo che l'errore di stima e la stima non sono correlati, la varianza del vero valore del parametro sarà semplicemente:

$$\text{var}(\theta) = \text{var}(\hat{\theta}) + \text{var}(\tilde{\theta}) \quad (3.33)$$

Trattando il parametro d'interesse come una variabile aleatoria, esso mostra una certa casualità che presenta due componenti: la componente relativa all'incertezza della stima, governata dal movimento che la stima effettua per avvicinarsi ai movimenti del vero valore del parametro, e l'incertezza residua che rimane anche dopo aver fatto la stima, la varianza dell'errore di stima. Per avere l'errore più piccolo possibile, la situazione ideale sarebbe avere una $\text{var}(\tilde{\theta})$ molto bassa in modo che tutta la varianza venga assorbita da $\hat{\theta}$ ¹⁸. La situazione migliore sarebbe quella in cui l'incertezza del vero valore del parametro venga assorbita completamente dall'incertezza della stima, con conseguente azzeramento della variabilità residua [Bertsekas and Tsitsiklis, 2008] [Tsitsiklis, 2010].

Stima LMS lineare. Tornando all'esempio precedente, si è visto che lo stimatore ottimale non era lineare. Questo fattore è però associato a maggiori elementi di complessità e, quando vi sono molteplici variabili aleatorie in gioco, i calcoli e gli integrali alla base del calcolo del valore atteso condizionato possono rivelarsi molto articolati.

¹⁸La varianza di $\tilde{\theta}$ sarebbe quindi molto elevata; si noti che la varianza dell'errore di stima e della stima sono inversamente proporzionali.

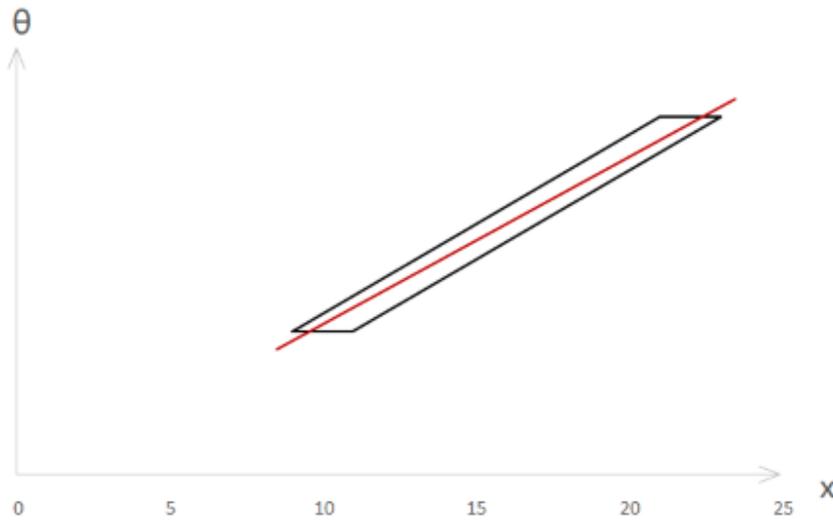


Figura 3.8: Esempio stimatore LMS lineare.

Uno strumento volto alla diminuzione del grado di complessità in casi analoghi all'esempio visto in precedenza, è rappresentato dagli stimatori LMS lineari. La logica di fondo di questa metodologia è vincolare gli stimatori a essere lineari, a essere funzioni lineari dei dati X , assumendo una forma di questo genere:

$$\hat{\theta} = aX + b \quad (3.34)$$

Il relativo errore quadratico medio sarà:

$$E[(\hat{\theta} - \theta)^2] = E[(aX + b - \theta)^2] \quad (3.35)$$

Minimizzando questa quantità per ogni a e per ogni b , si otterrà il *best linear estimator*, lo stimatore ottimo, ovvero:

$$\hat{\theta}_L = E[\theta] + \frac{Cov(X, \theta)}{var(X)}(X - E[X]) \quad (3.36)$$

Questo stimatore assumerà la forma descritta dalla linea rossa nella figura 3.8: si noti che esso è lineare e, in questo caso, descrive una covarianza positiva tra X e θ tramite la pendenza positiva della retta.

La formula 3.36 può essere così interpretata: $E[\theta]$ è la stima in assenza di informazioni

sul parametro d'interesse, in assenza di osservazioni sul fenomeno d'interesse, mentre la parte restante della formula è una sorta di correzione di $E[\theta]$ resa possibile dalle nuove informazioni contenute in X . L'ammontare di questa correzione è determinato dalla covarianza tra le due variabili aleatorie X e θ . Se la covarianza fosse nulla, ottenere informazioni dai dati non sarebbe rilevante per un aggiornamento dello stimatore, quindi $\hat{\theta}_L = E[\theta]$: lo stimatore ottimo sarebbe dunque lo stesso del caso di assenza di informazioni.

Nel caso in cui la covarianza fosse diversa da zero, le nuove informazioni contenute in X saranno invece il *driver* per una migliore stima del parametro d'interesse rispetto alla stima in assenza di informazioni.

L'errore quadratico medio relativo a $\hat{\theta}_L$ sarà:

$$E[(\hat{\theta}_L - \theta)^2] = (1 - \rho^2)\sigma_\theta^2 \quad (3.37)$$

Dalla formula si evince che questa quantità è influenzata da σ_θ^2 , ovvero dalla varianza della variabile originale: maggiore è la variabilità della V.A. originale, maggiore sarà l'errore quadratico medio dello stimatore lineare.

Si noti che questo fenomeno è mitigato dalla correlazione ρ tra i dati e i parametri d'interesse. Maggiore sarà l'indice di correlazione, minore sarà il termine $(1 - \rho)$ e di conseguenza minore sarà l'errore quadratico medio.

Con ρ pari a 1 l'errore quadratico medio sarebbe nullo, mentre se le due variabili aleatorie non fossero correlate, sarebbe pari alla varianza originale di θ ; le osservazioni non servirebbero a ridurre l'incertezza sul parametro d'interesse.

Quanto detto finora sulle stime LMS lineari è applicabile al caso in cui X e θ fossero dei vettori, delle quantità multidimensionali. Nel caso in cui avessimo una X vettoriale, lo stimatore vincolato a essere una funzione lineare dei dati sarebbe: $\hat{\theta} = a_1X_1, \dots, a_nX_n + b$. Si minimizzerebbe poi l'errore quadratico medio $E[(a_1X_1, \dots, a_nX_n + b - \theta)^2]$ impostando le derivate di $a_1X_1, \dots, a_nX_n + b$ rispetto ai coefficienti $(a_1, \dots, a_n; b)$ e ponendole pari a zero, si dovrà quindi risolvere un sistema di equazioni lineari soddisfatto da quei particolari coefficienti che minimizzino l'errore quadratico medio.

La cosa interessante che si scopre è che, al fine di trovare $\hat{\theta}_L$ nel caso di dati multipli, non serve conoscere l'intera distribuzione delle variabili aleatorie coinvolte ma basta conoscerne i valori attesi, le varianze e le covarianze in quanto, come già visto nella

formula 3.36, lo stimatore è completamente determinato quando si conoscono queste tre quantità.

Si pensi ora al seguente problema di stima: si fanno molteplici osservazioni rispetto a un certo parametro θ . Queste osservazioni sono disturbate e ϵ_i rappresenta gli errori, il rumore.

Si avrà quindi che $X_i = \theta + \epsilon_i$: le osservazioni non saranno esattamente rappresentative del parametro d'interesse reale a causa di una componente di errore. Si assume che gli errori siano indipendenti tra loro e che siano anche indipendenti dal vero valore del parametro. Si assume inoltre $\epsilon_i \sim (0, \sigma_i^2)$, gli errori in media sono zero e presentano delle varianze sconosciute σ_i^2 .

Il parametro d'interesse presenta una distribuzione a priori con una certa media e una certa varianza: $\theta \sim (\mu, \sigma_0^2)$.

Lo stimatore lineare ottimale sarà:

$$\hat{\theta}_L = \frac{\frac{\mu}{\sigma_0^2} + \sum_{i=1}^n \frac{X_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}} \quad (3.38)$$

Questa quantità è una media pesata e delle X e della media a priori μ : si noti che $\frac{1}{\sigma_i^2}$ è il peso dato alle X, $\frac{1}{\sigma_0^2}$ è il peso dato a μ e a denominatore vi è la somma di tutti i pesi.

Interessante è la forma dei pesi: ogni osservazione viene ponderata con l'inverso della varianza. Più l'osservazione è affetta da errore, minore sarà la sua utilità nella stima finale, per questo motivo le si assegna un peso minore.

Un altro aspetto interessante è che la media a priori viene trattata allo stesso modo di un'osservazione, si comporta come un'osservazione aggiuntiva.

L'inferenza bayesiana è quindi caratterizzata dal fatto che i parametri d'interesse sono modellati come variabili aleatorie, a differenza dell'inferenza classica in cui vengono trattati come delle costanti, e quindi vi è la necessità di assumere delle distribuzioni a priori. L'obiettivo è poi trovare la distribuzione a posteriori dopo aver fatto delle osservazioni sul fenomeno d'interesse e questo è il passaggio più complicato. I casi trattati sono stati appositamente quelli più semplici, ma nella realtà vi sono distribuzioni più complesse e multidimensionali; la formulazione della distribuzione a posteriori e il calcolo dei valori attesi condizionati possono risultare proibitivi dal punto di vista computazionale con tecniche ordinarie. Questo è il motivo della trattazione delle tec-

niche MCMC alla fine della prossima sezione: esse permetteranno di affrontare queste difficoltà e ottenere approssimazioni più che accettabili.

Infine, un altro aspetto delicato è la scelta della *prior*: essa è frutto di una decisione soggettiva del valutatore. L'esperienza e la conoscenza che un valutatore possiede su un determinato esperimento, su un certo argomento, sono sicuramente gli elementi alla base di una buona scelta della distribuzione a priori.

La difficoltà computazionale e la soggettività nella scelta della distribuzione a priori, sono stati a lungo elementi a favore degli statistici contrari a questo modo di approcciare l'inferenza.

Tuttavia, grazie al progresso tecnologico e all'idea sempre più diffusa tra gli accademici che anche nell'inferenza classica vi sia un forte livello di soggettività, questa metodologia ha trovato spazio e si è dimostrata essere uno strumento che si presta molto bene allo studio di fenomeni complessi difficilmente analizzabili con tecniche standard [Bertsekas and Tsitsiklis, 2008] [Tsitsiklis, 2010].

3.3 Tecniche di supporto all'inferenza bayesiana

3.3.1 Markov Chain

Le Catene di Markov sono una tipologia particolare di processi stocastici, ovvero di strumenti utili ad affrontare fenomeni che si evolvono nel tempo.

Negli esperimenti reali è molto raro avere a che fare con due variabili casuali, analizzarle e trarre subito le conclusioni. Molto più sovente, si parte con un certo numero di variabili aleatorie e con il tempo se ne generano di nuove, i processi stocastici servono proprio a catturare queste evoluzioni causate dal tempo. Le Catene di Markov sono essenzialmente una descrizione dell'evoluzione di un sistema di variabili in presenza di casualità, di rumore.

La decisione di cominciare la sezione approfondendo questo strumento, deriva dal fatto che le relative logiche sono alla base di potenti strumenti che hanno permesso di superare i limiti computazionali dell'inferenza bayesiana, limiti che, per molto tempo, hanno reso questa logica inferenziale difficilmente applicabile nello studio di fenomeni complessi.

Le Catene di Markov sono quindi composte da *stati* e, poiché stati casuali, si può parlare di variabili casuali: X_n rappresenta lo stato dopo “n” step temporali dal momento in cui il sistema ha iniziato a funzionare, X_0 .

Il passaggio da uno stato all’altro è determinato da una probabilità e graficamente è rappresentato da una freccia. Una semplice Catena di Markov potrebbe essere così rappresentata:

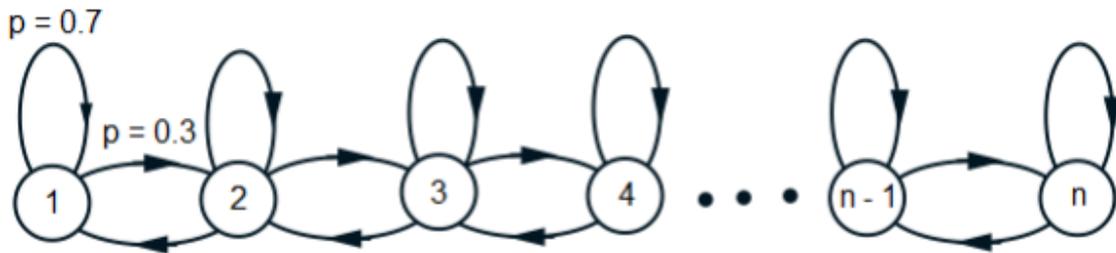


Figura 3.9: Esempio di una Catena di Markov.

Come si può notare, ci sono “n” stati e la probabilità di passare dallo stato 1 allo stato 2 è di 0,3 mentre la probabilità di rimanere nello stato 1 è di 0,7; ovviamente ogni freccia dovrà essere associata a una probabilità.

Questo processo si adatta sia per casi in cui il set dei possibili stati è finito, sia quando è infinito, sia per tempi discreti che per tempi continui. Il caso rappresentato nella figura è il caso migliore per iniziare a capire i concetti di questo strumento; ovvero il caso di un set di stati finito e il tempo discreto, così da riuscire visualizzare gli stati dopo 1, 2, 3, . . . , i, j transizioni, step temporali.

Il processo markoviano presenta la *proprietà di Markov*: ogni volta che il processo è allo stato i , la probabilità che allo step temporale successivo si trovi allo stato j , è un particolare valore $P_{i,j}$, non importa il percorso fatto per arrivare a i . Questa proprietà consente che la seguente uguaglianza sia vera:

$$P_{i,j} = P(X_{n+1} = j | X_n = i) = P(X_{n+1} = j | X_n = i, X_{n-1}, \dots, X_0) \quad (3.39)$$

Quindi, se il processo è nello stato i , non importa come il processo è arrivato a quel determinato punto, si ignorano completamente le transizioni precedenti e l’unica cosa

che si può affermare è che, dato che si è nello stato i , la probabilità di trovarsi nello step temporale nello stato j è pari a $P_{i,j}$. Questa proprietà indica che il passato non influenza gli eventi futuri.

Ovviamente, affinché questa proprietà sia applicabile, lo stato deve contenere tutte le informazioni necessarie per determinare di che tipologia saranno le transizioni future: ad esempio, sapere la posizione di un aereo in un determinato momento non è sufficiente, se ne conosciamo però anche la velocità, non importa dove fosse quell'aereo prima di quel momento, ma dal quel momento è possibile fare delle stime su dove l'aereo sarà in futuro, nell'attimo successivo.

Pertanto, quando si intende elaborare un processo markoviano, bisogna scegliere con cura quale sarà la variabile aleatoria di stato, si immaginano poi quali potrebbero essere i successivi stati e si identificano tutte le possibili transizioni. A questo punto si potrebbe decidere di escludere dal modello gli stati non utili al processo, ovvero gli stati che non presentano transizioni probabili da e verso altri stati.

Avendo una panoramica generale, bisogna ora capire effettivamente come si comporta il processo ma, essendo casuale, bisognerà accontentarsi di trovare le probabilità con le quali il processo si muoverà da uno stato all'altro.

Introduciamo dunque la *probabilità di transizione*, ovvero:

$$r_{i,j}(n) = P(X_n = j | X_0 = i) \quad (3.40)$$

Essa rappresenta la probabilità che, partendo da uno stato i , il processo si trovi allo stato j dopo “n” step temporali.

Prendendo un caso limite per farne comprendere il significato, si pensi a un processo in cui non vi sono transizioni; le probabilità saranno le seguenti:

- $r_{i,j}(n) = 1$ se $i = j$
- $r_{i,j}(n) = 0$ se $i \neq j$

Naturalmente se vi sono zero transizioni, allo step successivo il processo sarà esattamente dove ha avuto inizio e se quel determinato stato rappresenta anche il punto di arrivo, la probabilità di arrivare allo stato prefissato è pari a uno; se invece lo stato di

partenza e lo stato finale non coincidono, non ci sarà alcun modo per raggiungere lo stato finale, quindi la probabilità di transizione è nulla.

Siamo però interessati anche alle altre probabilità di transizione in “ n ” step temporali. Dato che “ n ” è un numero molto elevato, si utilizzerà la teoria della probabilità totale, condizionando su due diversi scenari, considerando i possibili modi in cui l’evento in questione può verificarsi.

Si è interessati dunque a capire qual è la probabilità di ritrovarsi allo stato j , “ n ” step temporali successivi all’inizio del processo, ovvero lo stato i al tempo zero.

Questo evento può verificarsi in molti modi diversi e si cercherà di raggrupparli in diversi sottogruppi. Una prima rappresentazione potrebbe essere quella descritta dalla figura 3.10.

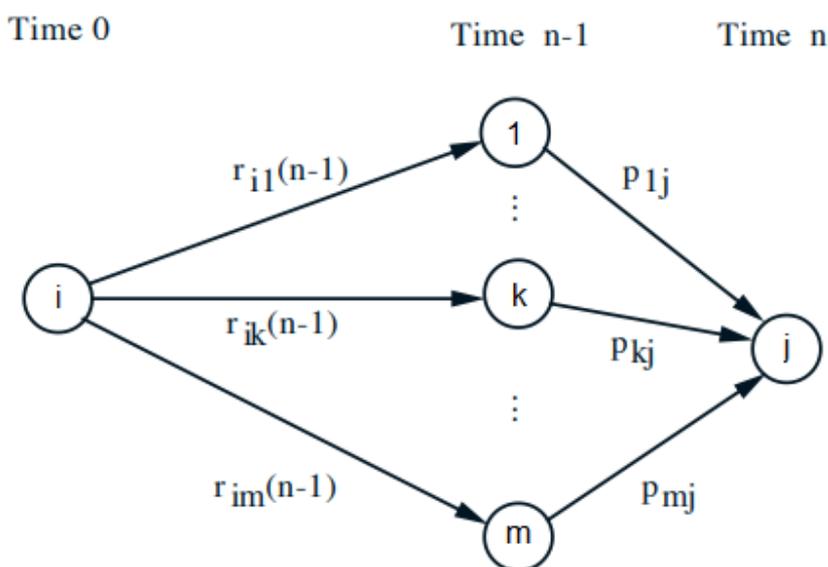


Figura 3.10: Scenario 1.

Come si può notare dall’immagine, si è inserita una fascia temporale intermedia corrispondente a $n - 1$ step temporali. Ad esempio, la freccia che parte da i e arriva a 1 corrisponde a moltissimi scenari diversi in cui si può verificare la transizione tra questi due stati e il termine $r_{i,1}(n - 1)$ corrisponde alla probabilità relativa all’insieme di tutti questi scenari.

Ci potrebbero essere però anche altre traiettorie nei primi $n - 1$ step temporali: è possibile che partendo da i si arrivi a m , e i modi in cui ciò può verificarsi sono molteplici; il tutto è catturato da $r_{i,m}(n - 1)$. Lo stesso ragionamento vale per la traiettoria che passa da k .

Dallo step $n - 1$ allo step K , partendo da 1, ci sarà una probabilità di arrivare a j pari a $p_{1,j}$, partendo da m , pari a $p_{m,j}$ e, partendo da k pari a $p_{k,j}$. Si può notare l'applicazione dell'assunto di Markov: le probabilità da $n - 1$ a n non dipendono dal percorso che il processo ha compiuto tra 0 e $n - 1$. Se questa proprietà non fosse applicabile, si sarebbero dovute individuare singolarmente tutte le traiettorie da 0 a $n - 1$ e poi si sarebbero dovute associare con le rispettive probabilità tra $n - 1$ e n : sarebbe stato tutto molto più complicato.

La probabilità totale di arrivare a j sarà, dunque, la somma delle probabilità dei due diversi scenari, ed è qui che si utilizza il teorema delle probabilità totali. Si sommeranno le probabilità relative al primo percorso, $r_{i,1}(n - 1)p_{1,j}$, relative al secondo percorso, $r_{i,k}(n - 1)p_{k,j}$ e al terzo, $r_{i,m}(n - 1)p_{m,j}$. Compattando quanto appena detto in una formula, si avrà:

$$r_{i,j}(n) = \sum_{k=1}^m r_{i,k}(n - 1)p_{k,j}, \forall i, j \quad (3.41)$$

Questa formula è una ricorsione applicata per ogni i, j simultaneamente, ovvero una tecnica che, per affrontare problemi complessi, li suddivide in sottoproblemi con caratteristiche molto simili ma più semplici. In questo caso, al posto di affrontare direttamente il problema relativo a “ n ” step di transizione, si calcola prima la probabilità relativa a $n - 1$ step e ciò permetterà di calcolare $r_{i,j}(n)$ più facilmente.

Ci sono anche delle varianti di questa formula. Se, per esempio, si fosse a conoscenza della probabilità dello stato iniziale e si volesse trovare la probabilità di essere allo stato j dopo “ n ” step temporali, la formula risulterebbe:

$$P(X_n = j) = \sum_{i=1}^m P(X_0 = i)r_{i,j}(n) \quad (3.42)$$

A questo punto, quello che interessa capire è se $r_{i,j}(n) \rightarrow r_j$, ovvero se $r_{i,j}(n)$ converga, dopo moltissimi step temporali, in qualcosa che abbia a che fare solamente con lo stato finale, rendendo ogni informazione sullo stato iniziale inutile. Per convergenza, si

intende che la probabilità $r_{i,j}$, con n che tende all'infinito, si stabilizzi su un determinato valore costante: la probabilità che il processo, dopo molto tempo, si stabilizzi su un certo stato è pari a quel determinato valore. Nel caso in cui la catena non convergesse, $r_{i,j}$ continuerebbe a oscillare tra i vari stati senza tendere a nessun valore.

Di solito le catene che non convergono sono caratterizzate da periodicità, cioè presentano un comportamento che si ripete. Un esempio di catena periodica è la seguente:

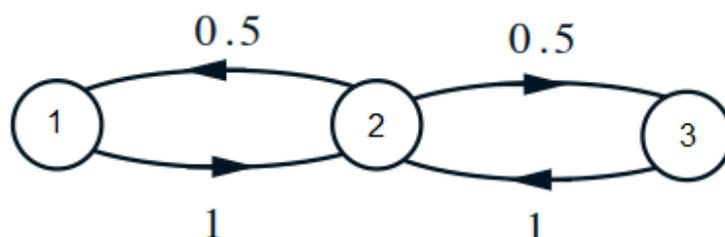


Figura 3.11: Esempio catena periodica.

In questa catena vi è forte periodicità: partendo dallo stato 2 e se il numero di step è pari, allora si è certi di tornare allo stato due e avremo che $r_{2,2}(n) = 1$. Se n è dispari, non ci sarà nessuna probabilità che il processo finisca nello stato 2 e avremo $r_{2,2}(n) = 0$. Con n che tende a infinito, le due probabilità relative a step temporali pari e a step temporali dispari non convergeranno ad alcun valore, ma oscilleranno continuamente tra 0 e 1.

Una catena per essere periodica richiede che, a partire da un cluster di stati, nello step successivo il processo cada in un cluster di stati differente. Se vi sono delle auto transizioni questa proprietà viene meno e quindi non sarà una catena periodica.

Un'altra cosa interessante da capire è se, nel caso di convergenza, ci possano essere catene in cui lo stato iniziale è rilevante. La risposta è affermativa: basti pensare a uno stato nel quale è solo possibile arrivare ma non uscire; in questo caso se si inizia in quel determinato stato non c'è nessuna probabilità di uscirne, quindi lo stato iniziale diventa di fondamentale importanza. Il caso appena descritto apre alla trattazione di un'importante classificazione riguardante gli stati o le classi di stati. Vi sono due tipologie di stati:

1. *Stati ricorrenti*: uno stato si dice ricorrente se, partendo da un certo stato iniziale i , c'è sempre un modo per ritornarci;
2. *Stati transitori*: uno stato si dice transitorio se, partendo da uno stato i , c'è la probabilità di arrivare in uno stato o in una classe di stati senza poter tornare al punto di partenza.

Se una catena presenta entrambe le categorie, lo stato di partenza sarà rilevante: se si parte da uno stato transitorio, nel lungo periodo il processo finirà sicuramente in uno stato ricorrente mentre se si parte da uno stato ricorrente, non ci sarà nessuna probabilità di finire in uno stato ricorrente.

Per fare in modo che lo stato iniziale non sia rilevante, la catena dovrebbe presentare solo una classe ricorrente, in modo che, in ogni caso, partendo sia dalla classe transitoria che dalla classe ricorrente, si finisca sempre nella stessa classe ricorrente [Bertsekas and Tsitsiklis, 2008] [Tsitsiklis, 2010] [Shiryaev, 2019].

Probabilità nello stato stazionario. L'obiettivo di questo paragrafo è analizzare il comportamento della catena in stato stazionario. Si vuole capire se $r_{i,j}$ converge a un valore di stato stazionario, indicato con π_j . Se il processo dura molto tempo, è possibile trovare la probabilità che lo il processo si stabilizzi su un particolare stato j . Abbiamo visto che non è sempre così, ad esempio nel caso di due classi ricorrenti.

Ci focalizzeremo sulle catene che presentano una sola classe ricorrente. In questo caso lo stato iniziale non ha alcun peso, condizione necessaria affinché la catena sia aperiodica e vi sia quindi la possibilità di ottenere la convergenza.

Da qui l'importante *teorema della convergenza allo stato stazionario*, il quale ha come assunto che vi deve essere una sola classe di stati ricorrenti e che quella classe ricorrente deve essere non periodica in modo che si possa parlare di *probabilità di stato stazionario*. Per calcolare questo tipo di probabilità, si parte da:

$$r_{i,j}(n) = \sum_{k=1}^m r_{i,k}(n-1)p_{k,j} \quad (3.43)$$

Si prende il limite da entrambe le parti e otteniamo:

$$\pi_j = \sum_{k=1}^m \pi_k p_{k,j}, \forall j \quad (3.44)$$

In effetti, il limite di n che tende a infinito della probabilità di transizione, il primo termine, è la probabilità di stato stazionario in j . Nel secondo termine, spostando il limite all'interno della sommatoria si ottiene che $r_{i,k}(n-1) = \pi_k$ poiché, quando n tende a infinito, la probabilità di transizione di k si trasforma nella probabilità di stato stazionario in k ; l'indice i scompare perché, come detto in precedenza, stiamo lavorando su una catena nella quale non ha importanza lo stato iniziale.

Si noti che non è una singola equazione, ma ci saranno tante equazioni, chiamate *equazioni bilanciate*, quanti sono gli stati finali j . Il sistema di equazioni generato, avrà un'unica soluzione se si impone questa naturale condizione quando si parla di probabilità:

$$\sum_j \pi_j = 1 \quad (3.45)$$

Questa *equazione di normalizzazione* impone che la somma delle probabilità degli stati j dovrà essere pari a uno. In questo modo avremo un sistema di equazioni tra le *equazioni bilanciate* e l'*equazione di normalizzazione* che ci permetterà di arrivare a una soluzione unica, ovvero la probabilità dello stato stazionario. Il sistema è quindi:

$$\begin{cases} \pi_j = \sum_{k=1}^m \pi_k p_{k,j} \\ \sum_j \pi_j = 1 \end{cases}$$

Molto importante è far passare il concetto che non è lo stato stesso a entrare nello stato stazionario, ma è la probabilità relativa agli stati stazionari che si stabilizza [Tsitsiklis, 2010].

3.3.2 Simulazione Monte Carlo

Questa tecnica nasce dall'intuizione del matematico Stanislaw Ulam che, annoiato durante un periodo di degenza, cercò di capire quale fosse la probabilità di vittoria a

solitario. Nonostante fosse un ottimo matematico, con i metodi combinatori classici fallì.

Pensò allora di usare un approccio frequentista, ovvero di contare il numero di vittorie su un numero enorme di partite giocate, ma, pensandoci bene, realizzò che non aveva tutto quel tempo a disposizione.

Cercando una soluzione, ebbe la brillante intuizione di far simulare il gioco a un computer e così nacque la simulazione di Monte Carlo.

La simulazione di Monte Carlo è un metodo per stimare i valori di una quantità sconosciuta usando la statistica inferenziale, ovvero utilizzare la statistica per generalizzare caratteristiche di un campione casuale sulla popolazione di riferimento, soddisfacendo determinate proprietà e avvalendosi di svariate tecniche.

Entrando nel vivo della simulazione Monte Carlo, si pensi di giocare alla roulette russa. Per iniziare si pensi a una roulette un po' semplificata, si può scommettere solo sui numeri, i quali vanno da 1 a 36. Il gioco è inoltre *fair*, corretto, ovvero è un gioco a somma zero e non c'è nessuna distorsione in favore del banco; alcune volte si vince e altre si perde, con un numero molto alto di scommesse ci si aspetterà una vincita pari a zero. Si può quindi scommettere su un numero compreso tra 1 e 36 e se si vince, si vince una somma pari a 35 volte la puntata mentre se si perde, si perde tutto. Scommettendo un euro, in caso di vittoria se ne ricevono 36, ovvero l'euro iniziale più i 35 vinti, se si perde, si perde tutto.

Il gioco inizia facendo girare la roulette, si lancia la pallina e la casella in cui cadrà la pallina è completamente casuale. Si scommette su un determinato numero, su una determinata casella e se cade sulla casella sulla quale si è scommesso si vince, altrimenti si perde e il gioco ricomincia.

Ora si immagini di far simulare questi passaggi da un software, facendo girare appositi codici sotto determinati vincoli: il tipo di gioco, il numero di giocate, la casella e l'importo oggetto di scommessa. Il simulatore di numeri casuali entra in azione e si vedrà poi se la scommessa è stata vincente oppure no. Inizialmente si guarderanno tre simulazioni di 100 giocate, un numero in linea con le giocate che potrebbe fare una persona in una nottata. Dalla simulazione di 100 giocate, quello che emerge è che il rendimento atteso è fortemente volatile. Il primo turno da 100 puntate potrebbe restituire un rendimento atteso pari al -100% , ovvero si perde tutto l'ammontare scommesso, il

secondo turno potrebbe restituire un valore atteso positivo anche abbastanza elevato, quindi una vincita corposa, mentre il terzo potrebbe generare di nuovo perdite.

Nel caso di turni da un milione di scommesse, situazione non associabile alle giocate di una persona, ma perfettamente in linea con i numeri che il Casinò è interessato ad analizzare, cosa accadrebbe ai rendimenti attesi?

Si può notare che i rendimenti attesi sono molto più vicini a zero e anche molto meno distanti tra loro, c'è molta meno varianza nei risultati. Più il numero di scommesse si avvicina a infinito, minore sarà la varianza dei rendimenti attesi, in quanto i loro valori saranno sempre più prossimi allo zero¹⁹.

Nella realtà, nei Casinò non si trovano roulette *fair* ma roulette con distorsioni a favore del banco.

Nel caso della roulette europea vi è un vantaggio del banco dato dalla casella verde corrispondente al numero zero. Se la pallina cade su questo numero, a meno che la puntata non sia stata proprio sullo zero, si perde l'intera somma scommessa; è quindi un metodo a favore del banco per diminuire i payoff dei giocatori. Nella roulette americana vi è un doppio vantaggio, determinato da uno zero e da un doppio zero.

Confrontando tra loro le simulazioni riguardanti infinite scommesse su una roulette corretta, su una europea e su una americana, si è notato che, nel caso di infinite giocate, nel primo caso il rendimento atteso si stabilizza intorno allo zero, nel secondo caso si ha un rendimento atteso negativo e nel terzo il rendimento atteso risulta essere ancora più sfavorevole al giocatore.

Nella realtà, nelle applicazioni dirette della probabilità, specialmente nei giochi d'azzardo, la Legge di Bernoulli e, in generale, le leggi che governano eventi casuali, spesso non vengono applicate nel modo corretto e si creano distorsioni.

Una distorsione tipica è la *fallacia del giocatore*: si intende un errore logico, un *bias* in genere dovuto all'euristica dell'ancoraggio che porta, ad esempio nei giochi d'azzardo, il giocatore ad ancorarsi a situazioni passate per prendere decisioni su situazioni future, anche se il gioco è completamente governato dal caso e gli eventi dovrebbero essere indipendenti.

Un esempio di questo *bias* è avvenuto proprio nel Casinò francese di Monte Carlo. Alla roulette, in una sessione di gioco, il colore nero è arrivato per ben 26 volte di fila.

¹⁹Questo fenomeno si rispecchia nella *legge di Bernoulli*, chiamata anche *Legge dei grandi numeri*.

Essendo il gioco completo e potendo scommettere sul colore, appena si sparse la voce, tutti i giocatori si precipitarono a scommettere sul rosso, in quanto ritenevano altamente improbabile un'ulteriore uscita del nero. Non avevano fatto i conti con l'indipendenza degli eventi governati da casualità; per questo motivo questa decisione collettiva giovò, e non poco, alle casse della casa da gioco.

La fallacia del giocatore non deve essere confusa però con la *regressione verso la media*, concetto ipotizzato da Galton alla fine del XIX secolo attraverso un'analisi compiuta sull'altezza dei figli rispetto a quella dei genitori. Nell'analizzare una quantità enorme di dati, osservò una cosa molto interessante: nel caso di genitori più alti della media, i figli risultavano essere più bassi dei genitori, mentre nel caso di genitori più bassi della media i figli risultavano più alti dei genitori. Questa evidenza trasmette il principio secondo il quale a seguito di un evento estremo, l'evento successivo sarà probabilmente meno estremo.

Per cogliere la differenza con la fallacia del giocatore, si torni all'esempio precedente relativo all'uscita di 26 neri di fila. In questo caso la fallacia del giocatore tende a far pensare che debba esserci una sorta di compensazione, ovvero che dovrebbero uscire molti più rossi per riequilibrare lo stato degli eventi. Nella logica della regressione verso la media, invece, si dirà che è probabile che nei prossimi 26 lanci si vedranno meno neri, cioè che l'evento sarà un po' meno estremo; ciò non significa che dovrebbero uscire meno di 13 neri per bilanciare l'evento estremo precedente, suggerirebbe un giocatore affetto da fallacia.

Con l'aumentare del numero di prove, più il campione è grande e più è probabile che ci si avvicini alla media, mentre con campioni più piccoli è più probabile osservare eventi estremi. Questo è il motivo per cui si chiama regressione verso la media ed è il fenomeno visto in precedenza con le simulazioni: più è ampio il campione, più la media si avvicinerà alla vera media.

Un metodo che permette di capire quanto un parametro osservato sia vicino al vero valore del parametro, è il metodo degli *intervalli di confidenza*: per stimare un valore sconosciuto è meglio non fermarsi solo alla media, ma creare un intervallo di confidenza che tenga conto di media e varianza.

Ipotizzando una confidenza del 95%²⁰, se si osserva ciò che accade per gli intervalli di

²⁰Il 95% dei dati campionari si trova a una distanza di 1.96σ dalla media, sotto l'ipotesi di errori

confidenza con le stesse simulazioni fatte in precedenza per la roulette corretta, quella europea e quella americana, si può notare che in ogni caso, con l'aumentare del numero di prove, l'intervallo di confidenza si riduce notevolmente. Ciò sta a significare che il range di oscillazione della media osservata si restringe e quindi c'è una buona probabilità che la media osservata sia vicina al vero valore del parametro [Guttag, 2016].

I metodi Monte Carlo si avvalgono quindi di tecniche di campionamento casuale e di simulazioni informatiche per ottenere soluzioni che con classici metodi di calcolo numerico sarebbero impossibili o molto complicati. Alla base di questi metodi vi è la generazione di variabili aleatorie, alle quali il ricercatore assegna specifiche proprietà; queste proprietà definiranno la popolazione di riferimento dei campioni oggetto di studio. Con l'*output* della simulazione, si potranno confrontare i risultati ottenuti dallo studio del campione con le proprietà definite per la popolazione e ricavare informazioni utili sul problema oggetto di analisi [Pastore, 2013].

L'esperimento relativo al gioco della roulette è un'applicazione semplice di questi metodi, ma è utile per comprenderne le logiche sottostanti.

La grande rivoluzione resa possibile da queste tecniche è associabile alla risoluzione di problemi che richiedono un'enorme capacità di calcolo, problemi non risolvibili con metodi analitici classici.

Nella prossima sezione vedremo una tecnica che unisce la simulazione di Monte Carlo e le Catene di Markov per risolvere problemi tipici dell'inferenza bayesiana, problemi legati alla risoluzione di integrali multidimensionali complessi.

3.3.3 Markov Chain Monte Carlo

La logica delle Catene di Markov, combinata al campionamento casuale ripetuto tipico delle simulazioni Monte Carlo, dà alla luce i metodi MCMC. Questa tecnica computazionale ha avuto un ruolo centrale per problemi complessi e, nel caso di problemi inferenziali, ha permesso l'avvento dell'inferenza bayesiana, in quanto permette di risolvere la formula alla base e che nella maggior parte dei casi risulta di difficile risoluzione.

Questo metodo permette di risolvere integrali multidimensionali che, con metodi computazionali classici, potrebbero risultare non risolvibili per l'enorme capacità di calcolo distorti, la cui media pari a 0, e distribuiti secondo una Normale.

colo richiesta. Grazie a questa peculiarità, i metodi MCMC permettono di trattare distribuzioni molto complesse caratterizzate da uno spazio dei parametri a più dimensioni²¹.

Vi è quindi una distribuzione di probabilità multidimensionale e si cerca di ricavarne una distribuzione a posteriori tramite le MCMC. L'idea che ne è alla base consiste sostanzialmente nell'effettuare una *random walk*, una procedura iterativa, finalizzata a campionare la distribuzione a posteriori in quelle regioni dello spazio dei parametri dove vi è una probabilità maggiore, tralasciando le aree a minore probabilità.

Si pensi di iniziare da un punto casuale della distribuzione e di scegliere poi, sempre casualmente, un punto ad esso vicino. Si valuta la probabilità del secondo valore e se questa è maggiore²² rispetto alla probabilità legata al primo, ci si sposta su questo nuovo punto e si ripete l'operazione, altrimenti si ripete l'operazione partendo semplicemente dal punto iniziale.

Eseguendo questa procedura infinite volte si riesce a coinvolgere ogni valore dello spazio dei parametri, della distribuzione di probabilità, con una frequenza proporzionale alle relative probabilità, consentendo così di plottare la *posterior* nel modo più accurato possibile.

Le Catene di Markov si concretizzano in una sequenza di punti, di eventi, quindi di variabili aleatorie, nello spazio campionario che sono descritti a livello grafico come nella figura 3.12.

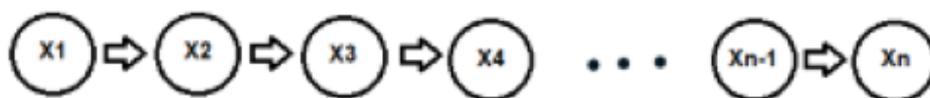


Figura 3.12: Esempio grafico Catena di Markov.

²¹Se il numero dei parametri, quindi di dimensioni, è maggiore di quattro, è l'unico metodo utile poiché utilizzare le griglie e i metodi utilizzati per i casi a due/tre dimensioni, diventa impossibile.

²²Ci sono anche delle funzioni che accettano valori con probabilità minori nel caso in cui presentino determinate caratteristiche, cioè ammettendo che passino determinati *test di accettazione*, in modo da poter far risultare nel plot della distribuzione anche gli spazi legati ai punti meno probabili.

La cosa interessante di questo modello è che ogni evento della sequenza dipende esclusivamente dall'evento precedente, quindi la probabilità a esso associabile è:

$$P(X_n|X_{n-1}, \dots, X_4, X_3, X_2, X_1) = P(X_n|X_{n-1}) \quad (3.46)$$

Ciò che succede prima di X_{n-1} , non rileva in alcun modo per la determinazione di X_n : come già visto nella sezione specifica sulle Catene di Markov, questa è una proprietà desiderabile, anche se si è visto che ci sono delle Catene in cui anche lo stato iniziale è rilevante.

Nel continuare la trattazione, i ragionamenti verranno fatti sulla base di catene descritte dall'equazione 3.46.

Si immagini una catena discreta di questo tipo:

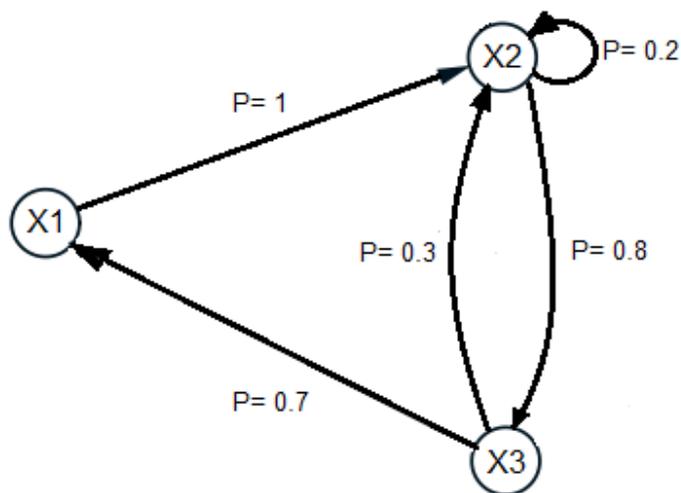


Figura 3.13: Esempio di catena discreta.

Essa è associabile alla seguente *matrice di transizione* T :

$$T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.2 & 0.8 \\ 0.7 & 0.3 & 0 \end{bmatrix} \quad (3.47)$$

La prima riga e la prima colonna di T sono relative a X_1 , la seconda riga e la seconda colonna a X_2 e l'ultima riga e l'ultima colonna a X_3 . Si immagini di partire da un punto casuale con le seguenti caratteristiche:

$$X_0 = (0,6; 0,3; 0,1) \quad (3.48)$$

Vi è una probabilità pari a 0,6 che lo stato sia in X_1 , una probabilità di 0,3 che si trovi in X_2 e una probabilità di 0,1 che si trovi in X_3 .

Moltiplicando il vettore per la matrice si otterrà:

$$X_1 = X_0 \cdot T = (0,07; 0,69; 0,24) \quad (3.49)$$

Quindi si avrà una probabilità del 7% di essere in X_1 , del 69% di essere in X_2 e del 24% di essere in X_3 ²³.

Ripetendo questa operazione infinite volte, si avrà una convergenza in un vettore X_∞ con le probabilità di stato stazionario di X_1 , X_2 , X_3 : non importa il punto in cui è iniziato il processo, dal momento che X_0 potrebbe essere qualsiasi cosa, il processo convergerà sempre a X_∞ .

Le MCMC devono essere *ergodiche*, cioè devono consentire che ogni punto nello spazio possa essere visitato proporzionalmente alle relative probabilità.

A tal fine la catena dovrà essere:

1. *Aperiodica*, ovvero non vi deve essere la probabilità di rimanere "intrappolati" in un determinato stato o classe di stati;
2. *Irriducibile*, ovvero ogni stato deve essere collegato a un altro stato con una probabilità maggiore di zero.

A questo punto si può parlare di MCMC e, applicandole al tema di questa trattazione, ovvero all'inferenza bayesiana, ci consentirà di mappare una distribuzione chiamata *posterior*, ovvero una distribuzione di probabilità congiunta di alcuni parametri d'interesse θ condizionata su determinate osservazioni vincolanti e su un certo modello-ipotesi .

²³Si noti che la somma di queste probabilità è pari a uno.

Le catene utili a tale scopo saranno dunque catene composte da campioni della *posterior* che convergono a una soluzione unica e di stato stazionario.

Un primo aspetto da tenere in considerazione è il periodo di *burn-in*, vale a dire quell'arco temporale antecedente al raggiungimento della distribuzione stazionaria, la *posterior*. I campioni relativi a questo periodo dovrebbero essere in genere scartati in quanto poco rappresentativi della distribuzione obiettivo.

Bisogna anche considerare la covarianza tra i parametri: se vi è molta covarianza, l'algoritmo è meno efficiente, mentre una covarianza bassa permette una convergenza più veloce.

Per costruire le catene MCMC ci si avvale di diversi algoritmi, i quali sono strutturati per rendere il processo il più efficiente possibile, in termini di minimizzazione del periodo di *burn-in* e di minimizzazione della covarianza tra i parametri²⁴ [Rossi et al., 2012] [Andrieu and Thoms, 2008] [Prügel-Bennett, 2020].

3.3.3.1 Algoritmo di Metropolis-Hastings

L'obiettivo di questo algoritmo è quello di generare una collezione di stati che segua, nel caso bayesiano, la distribuzione a posteriori.

L'algoritmo prende forma da un'ipotesi iniziale X^0 , procedendo poi con l'iterazione per i che va da 0 a $n - 1$ attraverso il seguente schema:

1. Campione: $u \sim U[0, 1]$, u è quindi un valore campionato da una distribuzione uniforme tra i valori 0 e 1;
2. Campione: $x^* \sim q(x^*|x^{(i)})$, dove $q(x^*|x^{(i)})$ viene chiamata *proposal distribution*, distribuzione proposta;
3. $u < A(x^{(i)}, x^*) = \min\{1, \frac{p(x^*) \cdot q(x^{(i)}|x^*)}{p(x^{(i)}) \cdot q(x^*|x^{(i)})}\}$, se questa condizione è soddisfatta ci si sposta dal punto in cui lo stato si trova al punto proposta x^* .

Si inizierà quindi campionando dalla *proposal*: partendo da un qualsiasi punto $x^{(i)}$ si effettuerà una *random walk* nei punti vicini, generando così punti proposta x^* .

²⁴Esistono delle tecniche che accettano campioni generati dopo una sequenza di numeri (per esempio uno ogni mille), in modo da diminuire al massimo il grado di correlazione e generare quindi campioni più indipendenti possibile.

La distribuzione proposta genera dunque un punto proposta, il quale dipende solamente dal punto precedente; da ciò si comprende perché gli algoritmi MCMC sono delle Catene di Markov.

Il software utilizzato per far girare l'algoritmo calcolerà il rapporto all'interno della funzione *min*, rapporto che a numeratore ha la probabilità del punto proposta moltiplicata per la distribuzione di probabilità del punto precedente dato il punto proposta mentre, a denominatore, presenta la probabilità del punto precedente moltiplicata per la distribuzione proposta, ovvero la probabilità di muoversi nel punto proposta dato il punto in cui si trovava il processo nello step temporale precedente.

Il rapporto è così strutturato per superare la restrizione che caratterizza l'algoritmo di Metropolis originale, ovvero funziona solamente per distribuzioni proposta simmetriche, vincolate dalla scelta di q in modo da rendere vera l'uguaglianza $q(x^*|x^{(i)}) = q(x^{(i)}|x^*)$ e semplificando il rapporto nella funzione *min* in $\frac{p(x^*)}{p(x^{(i)})}$.

La maggiore complessità del rapporto nella funzione *min* dell'algoritmo Metropolis-Hasting permette dunque anche la trattazione di distribuzioni asimmetriche, risultando quindi di maggiore applicabilità.

Tornando al punto 3, il significato della disuguaglianza è il seguente: si confronta un generico valore u campionato da una distribuzione uniforme con il minimo tra 1 e il risultato del rapporto. Ovviamente se il numeratore del rapporto è maggiore del denominatore e, quindi, il punto di proposta è più probabile del punto precedente, il rapporto sarà maggiore di uno e si sceglierà 1; essendo u minore di 1, la condizione del punto 3 sarà soddisfatta e si passerà dal punto precedente al punto proposta. Nel caso in cui il rapporto risultasse inferiore a uno, dovendo scegliere il minore dei valori nella funzione *min*, il software sceglierà il valore risultante dal rapporto. A questo punto si aprono due scenari: nel primo scenario il valore risultante dal rapporto risulta comunque maggiore di u e l'algoritmo accetta il punto di proposta²⁵, nel secondo scenario l'*output* del rapporto è inferiore di u , la condizione del punto 3 non è soddisfatta e quindi il processo ricomincia dal punto precedente²⁶.

Un aspetto importante è che le p nel rapporto non devono essere per forza essere le PDF complete ma basta avere una quantità che sia proporzionale alla funzione di densità complessiva. Si pensi a una distribuzione che presenta un denominatore, come nel caso

²⁵ $x^{i+1} = x^*$.

²⁶ $x^{i+1} = x^i$.

della distribuzione a posteriori bayesiana; si avrebbe che il rapporto tra $\frac{p(x^*|\theta) \cdot p(\theta)}{p(x)}$ ²⁷ e $\frac{p(x^i|\theta) \cdot p(\theta)}{p(x)}$ ²⁸, riscritto in termini di prodotto sarebbe:

$$\frac{p(x^*|\theta)p(\theta)}{p(x)} \cdot \frac{p(x)}{p(x^i|\theta) \cdot p(\theta)} \quad (3.50)$$

Si noti che il denominatore si annulla, ecco il motivo di quanto detto sopra: una funzione che sia proporzionale alla PDF totale è più che sufficiente.

Il metodo MCMC è dunque una tecnica molto utile quando si ha a che fare con funzioni molto complicate, a più dimensioni, e permettere di tracciare nuove distribuzioni e trovarne i relativi valori attesi tramite l'implementazione di appositi algoritmi.

L'algoritmo appena trattato è il metodo MCMC più semplice, ma è allo stesso tempo la base per la comprensione degli algoritmi più complessi [Roberts et al., 2001].

3.3.3.2 Algoritmo di Gibbs

Nell'algoritmo di Metropolis-Hastings vi è la necessità di specificare una *proposal distribution* per l'implementazione dello stesso.

Nel caso di distribuzioni multivariate a molte dimensioni, la suddetta specificazione può risultare di difficile determinazione in quanto le diverse dimensioni potrebbero presentare diverse peculiarità.

La struttura dell'algoritmo di Gibbs [Geman and Geman, 1984], per ottenere la distribuzione stazionaria di interesse, la *posterior*, è stata formulata per far fronte a questo problema prevedendo diversi campionamenti per le diverse dimensioni²⁹. L'idea di base è di campionare da distribuzioni condizionate univariate di ogni parametro casuale, dati gli altri parametri del modello [Locatelli et al., 2003]. Tutti i parametri tranne uno, quello su cui si condiziona, vengono considerati come costanti e vengono assegnati loro valori fissi.

Queste distribuzioni sono molto più semplici da simulare rispetto alle distribuzioni congiunte, in quanto generalmente presentano forme più semplici: si simulano dunque

²⁷Corrisponde a $p(x^*)$.

²⁸Corrisponde a $p(x^i)$.

²⁹Diversamente dall'algoritmo Metropolis-Hastings, il quale effettua campionamenti da una dimensione a un'altra senza restrizioni e cerca di campionare direttamente la distribuzione a posteriori congiunta.

“n ” variabili casuali sequenzialmente dalle “n ” distribuzioni univariate condizionate, anziché generare un singolo vettore n-dimensionale in un solo colpo dalla distribuzione congiunta.

Per iniziare, si espone il caso dell’analisi di una variabile aleatoria bivariata congiunta $\theta = (\theta_1, \theta_2)$. Le ipotesi alla base dell’algoritmo sono le seguenti:

- Si supponga $\theta_1, \theta_2 \sim p(\theta_1, \theta_2)$ e che sia possibile campionare dalle distribuzioni condizionate $p(\theta_1|\theta_2)$ e $p(\theta_2|\theta_1)$;
- Si inizia con un valore iniziale per θ_1 e θ_2 , ovvero $\theta_1^{(0)}$ e $\theta_2^{(0)}$.

L’algoritmo sarà così implementato:

1. $\theta_1^{(j)} \sim p(\theta_1|\theta_2^{(j-1)})$, ovvero all’iterazione j si campiona θ_1 dalla distribuzione di θ_1 condizionata sul valore di θ_2 all’iterazione precedente;
2. Si campiona poi un valore per θ_2 in questo modo: $\theta_2^{(j)} \sim p(\theta_2|\theta_1^{(j)})$. Alla j -esima iterazione si camperà un valore per θ_2 dalla distribuzione condizionata di θ_2 dato il valore di θ_1 appena campionato.

Utilizzando questa logica, il vettore $\theta^{(j)}$ composto da θ_1 e da θ_2 , converge in probabilità alla distribuzione congiunta d’interesse $p(\theta_1, \theta_2)$.

Inoltre la sequenza $\theta^{(j)}$ non è indipendente in quanto il relativo stato al tempo j dipende strettamente dallo stato assunto al tempo $j - 1$: si tratta dunque di una Catena di Markov [Carlo, 2004].

L’algoritmo, nel caso di una variabile N-dimensionale $\theta = (\theta_1, \dots, \theta_N)$, assumerà la seguente forma:

$$\theta_1^{(j)} \sim p(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_K^{(j-1)})$$

$$\theta_2^{(j)} \sim p(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_K^{(j-1)})$$

$$\theta_k^{(j)} \sim p(\theta_k | \theta_1^{(j)}, \dots, \theta_{k-1}^{(j)}, \theta_{k+1}^{(j-1)}, \dots, \theta_K^{(j-1)})$$

$$\theta_K^{(j)} \sim p(\theta_K | \theta_1^{(j)}, \dots, \theta_{k-1}^{(j)})$$

Queste distribuzioni sono chiamate *distribuzioni condizionate totali* in quanto si condiziona su tutti i parametri tranne che su quello che si cerca di plottare, di campionare. Nel caso appena descritto, i parametri $\theta_1, \dots, \theta_K$ della variabile multidimensionale θ sono stati considerati scalari. Tuttavia, l'algoritmo funziona anche nel caso di parametri a più dimensioni, fintanto che campionare dalle distribuzioni condizionate totali risulta possibile.

Quindi, per costruire questo strumento, è necessario essere in grado di plottare campioni dalle distribuzioni condizionate totali per la componente θ_K date tutte le altre diverse componenti [Neal, 1993].

4. Modello logit e sue estensioni

In questa sezione si affronterà a livello teorico lo strumento statistico che verrà utilizzato nel capitolo 5 per l'analisi sperimentale relativa alle preferenze elettorali. Tale strumento è aderente alle logiche sottostanti ai modelli di scelta discreta, ovvero quei modelli volti alla descrizione, alla comprensione e alla predizione delle scelte compiute dai *decision makers*, laddove il set di alternative disponibile sia composto da un numero finito di elementi. Si tratta dunque di uno strumento per lo studio di variabili dipendenti categoriali.

Per una più chiara esposizione del modello, la sezione si articolerà sviluppando il modello logistico multinomiale in chiave bayesiana partendo dalla sua genesi e introducendo gradualmente elementi di complessità.

4.1 Modello logit binomiale

La maggior parte dei fenomeni economico-sociali oggetto di analisi dei ricercatori hanno natura qualitativa. Molto spesso, infatti, l'oggetto di studio concerne aspetti del comportamento umano in risposta a determinati *input* di natura sia qualitativa che quantitativa.

A livello matematico, il comportamento umano è dunque la variabile dipendente, chiamata anche variabile risposta, mentre gli *input* che lo influenzano sono chiamati variabili indipendenti o covariate. È molto importante, in fase di formulazione dello studio, stabilire il nesso causale tra variabili dipendenti e indipendenti per non commettere un grave errore logico che inficerebbe i risultati dell'analisi.

Il modello logistico si presta dunque all'analisi dell'intensità degli effetti esercitati dalle covariate sulle variabili risposta, i quali, come si vedrà, non sono lineari come nella

classica regressione lineare.

Inoltre, il modello logit standard è finalizzato allo studio di variabili risposta dicotomiche, ovvero di variabili articolate in due categorie, molto frequenti nelle analisi dei processi di scelta: si pensi alla scelta di recarsi o meno alle urne in caso di elezioni, oppure alla scelta di posizione riguardante l'abrogazione o meno di una legge o un atto avente forza di legge tramite un *referendum* abrogativo.

Le variabili dicotomiche sono quindi sempre articolate in due categorie contrapposte che possono essere espresse in termini di "Sì" e "No", a seconda che il fenomeno oggetto di studio, la caratteristica in questione, sia presente o meno.

Il comportamento delle due categorie conviene però tradurlo in linguaggio binario, associando al "Sì", ovvero al caso in cui la caratteristica sia presente, il valore 1, mentre al "No" il valore 0. Si associa inoltre a 1 una probabilità pari a p e a 0 una probabilità pari a $1 - p$.

Nel caso di variabili dipendenti dicotomiche categoriali, l'obiettivo è, come già anticipato, la comprensione della misura in cui una o più covariate influenzino la probabilità di successo. Per probabilità di successo si intende la probabilità di osservare la caratteristica d'interesse nella variabile risposta¹.

A tal scopo, utilizzare modello di regressione lineare non sarebbe opportuno perché esso è strutturato per generare valori predetti della variabile dipendente che possono variare teoricamente da meno infinito a più infinito in maniera lineare e gli errori vengono considerati normali con varianza costante.

Nel caso di variabile dipendente dicotomica, bisogna invece restringere il *range* dei possibili valori assumibili dalla probabilità di successo in quanto, essendo una probabilità, può variare solamente all'interno dell'intervallo $[0; 1]$. Inoltre, l'effetto β esercitato da ogni regressore X sulla variabile dipendente Y non presenta natura lineare, gli errori non sono distribuiti normalmente e non presentano varianza costante.

Il modello di regressione logistica presenta le caratteristiche per catturare effetti non lineari e per generare valori predetti di Y compresi tra 0 e 1: queste peculiarità hanno però un costo in termini di maggiore complessità per la stima dei parametri del modello.

¹Per capire quanto influenzino la probabilità di insuccesso non serve un apposito studio ma basta ragionare per differenza.

La probabilità di successo $\pi(x_i)$ deve essere infatti trasformata in una sua particolare funzione chiamata *logit*. In prima battuta si definisce l'*odds*, ovvero il rapporto tra probabilità di successo e probabilità di insuccesso:

$$\text{odds} = \frac{\pi(x_i)}{1 - \pi(x_i)} = \frac{\pi(x_{i,1})}{\pi(x_{i,0})} = \frac{P(Y_i = 1|\mathbf{X}_i)}{P(Y_i = 0|\mathbf{X}_i)} \quad (4.1)$$

Si applica poi il logaritmo naturale all'*odds* e si ottiene:

$$\text{logit}(\pi(x_i)) = \ln(\text{odds}) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \ln\left(\frac{P(Y_i = 1|\mathbf{X}_i)}{P(Y_i = 0|\mathbf{X}_i)}\right) \quad (4.2)$$

Questa particolare funzione permette di trattare l'intercetta α , i parametri β_k e i regressori X_k in maniera lineare:

$$\text{logit}(\pi(x_i)) = \alpha + \sum_{k=1}^K \beta_k X_{i,k} \quad (4.3)$$

Il secondo termine dell'equazione, composto da una combinazione lineare di parametri e regressori, prende il nome di *predittore lineare* e si indicherà con L_i . All'interno del predittore lineare, l'intercetta α rappresenta il valore predetto della funzione logit della probabilità di successo quando tutti i regressori sono nulli, mentre ogni parametro β_k esprime la variazione di $\text{logit}(\pi(x_i))$ per ogni incremento unitario del relativo regressore X_k , trascurando gli effetti esercitati dagli altri regressori.

Il modello in forma additiva espresso dall'equazione 4.3 presenta, tuttavia, un'unità di misura della variabile risposta differente dall'unità di misura originaria, non garantisce ancora la non linearità degli effetti e che i valori predetti di Y siano compresi tra zero e uno, le due condizioni alla base della validità del modello.

Per ovviare a tali problemi, la soluzione è riformulare il modello in forma moltiplicativa, ovvero:

$$\pi(x_i) = P(Y_i = 1|\mathbf{X}_i) = \frac{e^{(\alpha + \sum_{k=1}^K \beta_k X_{i,k})}}{1 + e^{(\alpha + \sum_{k=1}^K \beta_k X_{i,k})}} \quad (4.4)$$

Grazie a questa formulazione, gli effetti esercitati dalle covariate sulla probabilità di successo risultano non lineari e le probabilità di successo sono vincolate a variare nell'intervallo $[0; 1]$.

La relazione tra i predittori lineari e le probabilità di successo assume una forma sigmoide, la quale suggerisce che, per valori molto bassi di L_i , un loro incremento comporta variazioni minime sulle probabilità predette $\pi(x_i)$. All'aumentare di L_i , anche l'effetto esercitato sulle probabilità predette tende ad aumentare ed è massimo in $\pi(x_i) = 0.5$. Oltre a questo valore, l'effetto in questione torna a diminuire fino a risultare quasi irrilevante.

La rappresentazione grafica della relazione tra predittori e probabilità predette fornita dalla figura 4.1 fornisce una più rapida intuizione rispetto alla spiegazione teorica appena fatta.

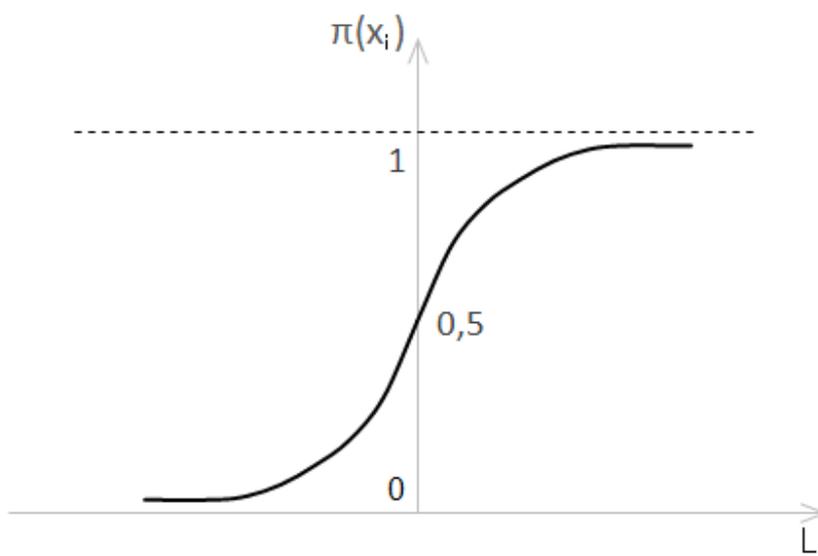


Figura 4.1: Regressione logistica.

Per quanto riguarda l'interpretazione dei risultati di una regressione logistica, bisogna porre attenzione sul fatto che gli effetti esercitati dai diversi regressori che la compongono sono lineari rispetto alla funzione logistica $\text{logit}(\pi(x_i))$, ma non sono lineari rispetto alla probabilità di successo predetta $\pi(x_i)$.

Per stimare gli effetti rispetto alla funzione logistica², basterà valutare le stime dei parametri β_k . Bisognerà invece porre attenzione sull'interpretazione dei parametri, in quanto l'unità di misura è espressa in unità *logit* e non in termini di probabilità di successo. Ciò comporta che dei parametri stimati l'informazione rilevante è il segno,

²Si analizzano i parametri stimati nella logica del modello additivo descritto dalla relazione 4.3.

il quale indica la variazione della probabilità di successo al crescere dei valori della relativa covariata³. Per determinare l'intensità degli effetti, un'analisi di questo tipo non è informativa.

Un metodo utilizzato per quantificare gli effetti sfrutta le caratteristiche del modello moltiplicativo valutando gli effetti generati da una variabile dipendente sulla probabilità di successo. Il metodo consiste nel confrontare le probabilità predette relative a diverse combinazioni delle variabili dipendenti, evidenziando la misura in cui gli effetti oggetti di studio possono variare. Nell'implementazione di questo metodo è importante tenere costanti i valori assunti dalle variabili di controllo, ovvero di quelle variabili supplementari che permettono di evitare una covariazione spuria tra X e Y; per *covariazione spuria* si intende quel fenomeno che spinge la covariazione osservata in assenza di altre variabili a non manifestarsi dopo che nuove variabili che possono plasmare l'iniziale relazione sono state introdotte [Corbetta et al., 2001].

Per la stima dei parametri, gli assunti del metodo dei minimi quadrati non sono rispettati. Utilizzando una logica inferenziale classica, la stima dei parametri si ottiene tramite il metodo della massima verosimiglianza/log-verosimiglianza⁴. Questo metodo, partendo da un determinato campione osservato per la variabile aleatoria d'interesse, consente di stimare i parametri grazie a stimatori che massimizzano la funzione di verosimiglianza o, equivalentemente⁵, la funzione di log-verosimiglianza e che presentano determinate desiderabili proprietà finite⁶ e asintotiche⁷ [De Bettin, 2014].

La stima dei parametri nell'ottica classica si concentra quindi sulla funzione di verosimiglianza $P_{X|\theta}(x|\theta)$.

Riprendendo però la formula 3.10, ovvero la formula di Bayes in termini di distribuzioni di probabilità, si può notare che la funzione di verosimiglianza ne è solo una parte, cioè il primo termine del prodotto al numeratore.

³Il segno positivo è indice di un aumento della probabilità di successo mentre il segno negativo di una sua diminuzione.

⁴La log-verosimiglianza è una semplice trasformazione della verosimiglianza in forma logaritmica, con lo scopo di semplificarne l'algebra.

⁵Massimizzare la verosimiglianza è equivalente a massimizzare la log-verosimiglianza in quanto quest'ultima è monotona crescente.

⁶Distorsione ed efficienza in termini di MSE, cioè *Mean Squared Error*, ovvero errore quadratico medio.

⁷Distorsione asintotica, consistenza in senso forte e consistenza in senso debole.

Nell'ottica inferenziale bayesiana, come ampiamente discusso nelle sezioni precedenti, non ci si limita agli *input* generati dalle stime di massima verosimiglianza, ma, assumendo che i parametri siano variabili aleatorie, e come tali governati da distribuzioni a priori assegnate dal ricercatore, si cerca di trovare una distribuzione di probabilità a posteriori dei parametri, ovvero dati i dati osservati. Si tratta dunque di un'estensione del metodo della massima verosimiglianza in quanto sfrutta pienamente la formula di Bayes.

Dato che nel modello logistico si ha a che fare con i stime di parametri discreti, si procederà con la logica relative ai problemi del test d'ipotesi, sia che i dati osservati siano discreti che continui.

Una volta computata la distribuzione a posteriori del parametro, si procede con il metodo M.A.P., illustrato nel paragrafo 3.2, per trovare il valore relativo al parametro più probabile della funzione obiettivo.

4.2 Modello logit multinomiale

Il modello logit multinomiale è strutturato per i casi in cui la variabile dipendente si articola in più di due categorie, ovvero non è dicotomica ma politomica. Si tratta dunque un'estensione del modello logistico appena trattato. Modelli di questo tipo sono adatti ad analizzare le scelte e le preferenze degli individui, motivo per cui è stato scelto in questa tesi per analizzare le preferenze elettorali.

L'obiettivo è stimare gli effetti causali che una o più covariate esercitano sulla propensione a scegliere tra più di due possibili scelte, la quale è espressa in termini di probabilità e quindi può solamente assumere valori all'interno dell'intervallo $[0; 1]$.

Gli elementi di complessità che caratterizzano questo modello sono causati dal numero di categorie in cui la variabile risposta si articola e si traducono in una più ardua stima e interpretazione dei parametri d'interesse.

Per iniziare con la spiegazione del modello, può essere utile evidenziare il collegamento con il modello logit binomiale: il modello logit multinomiale può essere trattato come la combinazione di due o più modelli logit binomiali.

Prima di generalizzare la forma del modello, si pensi al caso più semplice in cui la variabile dipendente si articola in tre categorie. Anche in questo caso si lavora sulla

funzione *logit* della probabilità di successo, ma la relazione 4.2 vista in precedenza non è più valida in quanto le probabilità da mettere a confronto non sono più solamente due ma sono tre.

Si dovranno confrontare la probabilità che Y assuma il valore 3 contro la probabilità che assuma il valore 1, la probabilità che Y assuma il valore 2 contro la probabilità che assuma il valore 1 e la probabilità che Y assuma il valore 3 contro la probabilità che assuma il valore 2. Non servono altri confronti poiché speculari a quelli citati: è possibile fare una prima generalizzazione, ovvero che per variabili che si articolano in J categorie, i confronti rilevanti fra coppie di categorie saranno pari a $J \cdot (J - 1)/2$.

Quindi essendo tre categorie avremo tre confronti, ovvero:

$$\ln \left(\frac{\pi(x_{i,3})}{\pi(x_{i,1})} \right) = \frac{P(Y_i = 3 | \mathbf{X}_i)}{P(Y_i = 1 | \mathbf{X}_i)} \quad (4.5)$$

$$\ln \left(\frac{\pi(x_{i,2})}{\pi(x_{i,1})} \right) = \frac{P(Y_i = 2 | \mathbf{X}_i)}{P(Y_i = 1 | \mathbf{X}_i)} \quad (4.6)$$

$$\ln \left(\frac{\pi(x_{i,3})}{\pi(x_{i,2})} \right) = \frac{P(Y_i = 3 | \mathbf{X}_i)}{P(Y_i = 2 | \mathbf{X}_i)} \quad (4.7)$$

Quindi, nel caso di tre categorie, si può ragionare in termini di tre modelli logistici binomiali. Tuttavia, è sufficiente considerare solo due confronti: ogni confronto può essere espresso in funzione degli altri due, dunque uno dei tre confronti risulterebbe superfluo. Si ha dunque una seconda generalizzazione per cui, nel caso di J categorie, ci saranno $J - 1$ possibili confronti indipendenti fra coppie di categorie.

Non vi è una regola specifica sulla scelta dei confronti, ma si segue la seguente logica: come prima cosa si sceglie la categoria di riferimento della variabile risposta e poi si pongono a confronto le altre categorie con quella di riferimento.

Per esempio, nel caso appena visto, se si sceglie la categoria 1 come categoria di riferimento, si dovranno stimare questi due modelli logit binomiali:

$$\text{logit}(\pi(x_{i,3})) = \ln \left(\frac{\pi(x_{i,3})}{\pi(x_{i,1})} \right) = \alpha_3 + \sum_{k=1}^K \beta_{k,3} X_{i,k} \quad (4.8)$$

$$\text{logit}(\pi(x_{i,2})) = \ln \left(\frac{\pi(x_{i,2})}{\pi(x_{i,1})} \right) = \alpha_2 + \sum_{k=1}^K \beta_{k,2} X_{i,k} \quad (4.9)$$

Si noti che l'unica differenza con il modello additivo 4.3 è rappresentata dal fatto che i parametri α e β presentino pedici rappresentativi della categoria di Y posta a confronto con la categoria di riferimento.

A questo punto è possibile, fissando la categoria 1 come quella di riferimento, esprimere la forma generale del modello logit multinomiale in forma additiva, ovvero:

$$\text{logit}(\pi(x_{i,j})) = \ln \left(\frac{\pi(x_{i,j})}{\pi(x_{i,1})} \right) = \alpha_j + \sum_{k=1}^K \beta_{k,j} X_{i,k} \quad (4.10)$$

L'indice j relativo alle categorie messe a confronto con quella di riferimento, assumerà valori compresi tra 2 e J , in quanto il valore $j = 1$ è quello relativo alla categoria di riferimento.

Ciò comporta che, nel caso di variabile riposta articolata in J categorie, vi saranno $J - 1$ diverse equazioni da risolvere simultaneamente *Non bisogna considerare i $J - 1$ modelli logit binomiale come se fossero indipendenti.* per stimare $J - 1$ insiemi diversi di parametri.

Il modello logistico multinomiale in forma moltiplicativa, che esprime la relazione tra le probabilità di successo predette $\pi(x_{i,j})$ e i $J - 1$ predittori lineari $L_{i,j}$, per $m = 2, \dots, J$ ⁸, assumerà la seguente forma:

$$\pi(x_{i,m}) = P(Y_i = m | \mathbf{X}_i) = \frac{e^{(\alpha_m + \sum_{k=1}^K \beta_{k,m} X_{i,k})}}{1 + \sum_{j=2}^J e^{(\alpha_j + \sum_{k=1}^K \beta_{k,j} X_{i,k})}} \quad (4.11)$$

Questo modello tiene conto del fatto che tutte le $J - 1$ equazioni stimate hanno come comune denominatore la stessa probabilità, ovvero la probabilità della categoria di riferimento $\pi(x_{i,1})$ e vale solamente quando la categoria della variabile dipendente è diversa dalla categoria di riferimento.

La probabilità che la variabile riposta assuma una categoria, rappresentata dal valore m , diversa da quella di riferimento in corrispondenza del soggetto i è, come nel caso del modello logit binomiale, funzione non lineare di tutti i parametri stimati.

⁸Sono le categorie diverse da quella di riferimento.

La probabilità che Y assuma la categoria di riferimento in corrispondenza del soggetto i sarà invece:

$$\pi(x_{i,1}) = P(Y_i = 1 | \mathbf{X}_i) = \frac{1}{1 + \sum_{j=2}^J e^{(\alpha_j + \sum_{k=1}^K \beta_{k,j} X_{i,k})}} \quad (4.12)$$

Infine, per quanto riguarda l'interpretazione dei parametri in termini *logit*, la logica è esattamente uguale a quella relativa al modello binomiale, con l'unica differenza che bisogna fare riferimento esplicito a quale categoria di Y viene posta a confronto.

Anche per quanto riguarda l'interpretazione in termini di probabilità di successo, la logica è la stessa del caso in cui la Y sia binomiale; essendo i parametri non lineari, si procederà con il calcolo delle probabilità predette relative a diverse combinazioni di valori delle variabili risposta, per poi confrontare i risultati ottenuti [Corbetta et al., 2001].

Inoltre, anche la stima dei parametri non differisce dal modello standard. Essa può seguire logiche inferenziali classiche o bayesiane e differisce dal modello binomiale solamente per una maggiore complessità computazionale richiesta per le stime.

4.3 Modello logit multinomiale misto in chiave bayesiana per modelli a scelta discreta

In questa sezione, l'obiettivo è avvicinarsi ulteriormente al modello che verrà implementato nel capitolo 5. Si cercherà di approfondire il modello logit multinomiale visto in precedenza, aumentare il livello di complessità introducendo nuovi concetti, e contestualizzarlo per processi di scelta influenzati dall'utilità che i *decision makers* assegnano a una determinata opzione scelta all'interno di un set finito di alternative discrete mutualmente escludibili.

Poiché l'oggetto di studio in questa tesi è il voto politico, si esporrà il caso in cui i *decision makers* sono gli elettori e le possibili scelte i vari partiti che si candidano alle elezioni. Si assume quindi che la scelta degli elettori per un determinato partito sia influenzata dall'utilità percepita nel votare quel partito piuttosto che votarne un altro. La massimizzazione dell'utilità percepita dall'elettore nel votare un determinato partito è dunque l'elemento chiave alla base dei differenti risultati elettorali. I politici hanno,

quindi, come obiettivo questa massimizzazione e i possibili *driver* che ne influenzano la riuscita o il fallimento, potrebbero avere, per esempio, natura ideologica oppure potrebbero avere a che fare con promesse o programmi elettorali e con aspetti legati alla comunicazione politica.

Gli elettori cercheranno dunque di massimizzare la loro utilità e i partiti lavoreranno per incanalare il maggior numero di voti, attraverso strategie politiche volte a massimizzare l'utilità per il maggior numero di elettori; tali strategie potrebbero essere direzionate da analisi statistiche volte a comprendere l'eterogeneità che caratterizza le decisioni di voto per i *decision-makers*.

Tra i diversi strumenti statistici, il modello logit multinomiale misto in chiave bayesiana si presta molto bene a catturare l'eterogeneità dei *decision-makers* e stimare, con un elevato grado di accuratezza, gli effetti generati dalle covariate: questo modello permette di assegnare ai coefficienti delle variabili introdotte nel modello valori diversi per diversi elettori e permette di rilassare l'assunzione dell'indipendenza delle alternative irrilevanti (IIA), operazioni non permesse in un modello multinomiale classico [Nava et al., 2016].

Si vuole rilassare l'assunzione IIA poiché, nel caso di elezioni politiche, potrebbe essere violata. Essa infatti implica che la probabilità di scelta non sia influenzabile dalla presenza o dall'assenza di ogni alternativa addizionale presente nel set di scelte [Solgaard and Hansen, 2003] e che gli attributi delle possibili scelte pesino tutti allo stesso modo [Fiebig et al., 2010].

Nel caso di voto politico, questa assunzione è facilmente violabile in quanto, come riportato anche in [Yamamoto, 2011], il comportamento elettorale è fortemente influenzato da variazioni nel set di alternative: si pensi a come semplici coalizioni politiche⁹, l'ingresso in politica di un nuovo partito¹⁰ oppure una variazione radicale nei valori di un partito¹¹ possano far variare, anche in maniera molto significativa, la decisione di votare un determinato partito rispetto alla decisione di votarne un altro¹².

Per impedire che l'assunzione IIA si verifichi, bisogna rilassare un'altra assunzione del modello multinomiale standard secondo la quale la componente stocastica dell'utilità

⁹Questo rappresenta un possibile caso di restrizione delle possibili alternative.

¹⁰Questo rappresenta un possibile caso di incremento delle possibili alternative.

¹¹Questo rappresenta un mutamento riguardante un attributo della scelta.

¹²Si desume dunque che, al variare delle caratteristiche del set di scelte, varino anche i processi decisionali dei *decision-makers* e che agli attributi che caratterizzano le varie alternative vengano attribuiti pesi diversi dei decisori.

si distribuisce indipendentemente e identicamente (IID) secondo una distribuzione dei valori estremi.

Assumere che gli errori siano IID, comporta assumere che, a seguito di un campionamento, elettori diversi che presentano le stesse caratteristiche¹³ abbiano le stesse preferenze e valutino allo stesso modo le caratteristiche di un certo partito. Questo comporta assumere che i coefficienti delle covariate incluse nel modello siano gli stessi per tutti gli elettori.

Ovviamente è un'assunzione che nella realtà è alquanto restrittiva e facilmente violabile ogni volta che vi siano dei fattori non osservabili dal valutatore che influenzano i processi di scelta del *decision-maker* [Solgaard and Hansen, 2003].

L'obiettivo del modello presentato in questa sezione è rilassare queste due assunzioni per analizzare e generare stime accurate nel caso di fenomeni governati in parte da variabili latenti. Il modello si presta alla scomposizione della componente stocastica in un elemento distribuito IID secondo una distribuzione dei valori estremi e un elemento che segue una distribuzione G che descrive l'eterogeneità tra gli elettori. Come affermato in [Sottile, 2015], questa scomposizione, in particolare l'assunzione di una distribuzione G che cattura l'eterogeneità dei decisori, è fondamentale per modellizzare, tra gli altri, casi caratterizzati da eterogeneità casuale nelle preferenze, correlazione tra le alternative e eteroschedasticità, ovvero quei casi che i modelli di regressione classici non sarebbero in grado di modellizzare correttamente.

Dopo aver appreso la rilevanza dell'utilità nei modelli a scelta discreta e averne anticipato alcuni aspetti, si passa a una traduzione quantitativa di questo concetto.

La probabilità che l'elettore i scelga il partito j in una determinata tornata elettorale può essere specificata in termini di utilità, con $k \in J^{14}$ e con $k \neq j$ si avrà che:

$$\pi_{i,j} = P\{U_{i,j} > U_{i,k}\} \quad (4.13)$$

La probabilità che l' i -esimo elettore scelga il partito j è pari alla probabilità che l'utilità ricevuta da quel partito sia maggiore dell'utilità ricevuta da un generico partito k [Solgaard and Hansen, 2003].

¹³A livello di reddito, di istruzione e di età per esempio.

¹⁴Con J si indica l'insieme che comprende tutti i partiti che si presentano alle elezioni.

L'utilità è composta da due componenti, una deterministica e una stocastica. L'utilità che l'elettore i ha nel scegliere il partito j può essere così scomposta:

$$U_{i,j} = V_{i,j} + \epsilon_{i,j} \quad (4.14)$$

$V_{i,j}$ è la componente deterministica dell'utilità, mentre $\epsilon_{i,j}$ è la componente stocastica, ovvero una componente inosservata e indipendente dalla componente deterministica, la quale è assunta, come anticipato, essere in parte costituita da due elementi non correlati. L'equazione 4.14 diventa:

$$U_{i,j} = V_{i,j} + \epsilon_{i,j} = \boldsymbol{\beta} \mathbf{X}_{i,j} + \eta_{i,j} + \epsilon_{i,j} \quad (4.15)$$

Il primo elemento della componente stocastica, $\epsilon_{i,j}$, è distribuito IID secondo la distribuzione dei valori estremi mentre il secondo elemento, $\eta_{i,j}$, segue una distribuzione G^{15} [Hensher and Greene, 2003]. A seconda dell'attributo associato a $\eta_{i,j}$ il modello generale si scompone in due varianti, ovvero nel *modello con parametri aleatori* e nel *modello con componente d'errore* [Sottile, 2015].

Nel primo modello, gli attributi associati al parametro random sono noti al valutatore e l'equazione 4.15 si trasforma in:

$$U_{i,j} = V_{i,j} + \epsilon_{i,j} = \boldsymbol{\beta}_i \mathbf{X}_{i,j} + [\eta_{i,j} + \epsilon_{i,j}] = \sum \beta_{i,l} X_{i,l,j} + \epsilon_{i,j} \quad (4.16)$$

Si noti che la componente deterministica dell'utilità è, come nel modello 4.15, composta da $\mathbf{X}_{i,j}$, ovvero dal vettore che rappresenta le caratteristiche osservate dell' i -esimo elettore in corrispondenza del j -esimo partito. Questa nuova specificazione, differisce però nel trattare il parametro relativo alle preferenze $\boldsymbol{\beta}$: esso non viene più considerato fisso ma, inserendo a pedice la i , si vuole indicare che varia a seconda dell'elettore preso in considerazione. Si può affermare che $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \eta_i$, ovvero che la preferenza di un determinato elettore sarà pari alla media elettorale, l'effetto fisso, più la componente stocastica governata da G relativa a quel determinato elettore.

La formula 4.16 assumerà quindi questa forma:

$$U_{i,j} = V_{i,j} + \epsilon_{i,j} = \boldsymbol{\beta} \mathbf{X}_{i,j} + [\eta_i \mathbf{X}_{i,j} + \epsilon_{i,j}] \quad (4.17)$$

¹⁵Quindi $\epsilon_{i,j} = \eta_{i,j} + \epsilon_{i,j}$.

È interessante notare che la parte di componente stocastica governata da G moltiplichi $\mathbf{X}_{i,j}$, determinando una correlazione tra le alternative all'interno della componente di residuo aleatorio.

Nel secondo modello, invece, gli attributi dei parametri random sono sconosciuti al valutatore e la 4.15 si trasforma in:

$$U_{i,j} = V_{i,j} + \epsilon_{i,j} = \beta_i \mathbf{X}_{i,j} + \eta_{i,j} + \epsilon_{i,j} \quad (4.18)$$

In questa sede ci si concentrerà sul primo modello definito dalla formula 4.17 e si procede con specificare la probabilità che l'elettore i selezioni un partito j durante una tornata elettorale, in quanto non si è a conoscenza dell'esatta utilità che viene assegnata da un certo elettore a un determinato partito.

Come detto in precedenza, questo modello si utilizza quando sono note le specifiche preferenze dei soggetti campionati, degli elettori in questo caso, e questo permette di calcolare la probabilità che un elettore scelga un determinato partito a seconda delle specifiche preferenze e, siccome una parte della componente aleatoria è assunta distribuita IID secondo una distribuzione dei valori estremi, si può esprimere in termini logit e, nella forma moltiplicativa, è così espressa:

$$\pi_i(j|\beta_i) = \frac{e^{(\sum_{l=1}^r \beta_{i,l} X_{i,l,j})}}{\sum_{k=1}^J e^{(\sum_{h=1}^r \beta_{i,h} X_{i,h,k})}} \quad (4.19)$$

A questo punto, dopo aver introdotto la logica dei *mixed logit* e specificato la versione utilizzata, si può introdurre come viene utilizzata la logica bayesiana per la stima dei β_i .

L'algoritmo di Gibbs è lo strumento che si presta meglio a tal fine, in quanto adatto alla stima di modelli stabili a scelta individuale. Come si è visto, questo strumento applica la logica del campionamento ripetuto e, in questo caso, campiona molte volte dalla *posterior* di \mathbf{B}_i ¹⁶ per trarre informazioni utili sul vettore di parametri; per comunicare quanto appreso sul vettore delle preferenze dallo studio dei dati, è utile sintetizzare quanto ottenuto dai campionamenti tramite apposite misure.

¹⁶Ovvero dalla distribuzione dei parametri una volta osservati i dati.

Entrando nel cuore della stima bayesiana, si utilizza un *modello gerarchico bayesiano* il quale, anche se il nome potrebbe essere ingannevole, non è un nuovo modello ma solamente una procedura di stima. Gli effetti casuali β_i , in assenza di conoscenze a priori, vengono assunti come IID secondo una distribuzione Normale multivariata r-dimensionale con i parametri vettore media $\bar{\beta}$ e matrice di covarianza Σ , ovvero $\beta_i \stackrel{IID}{\sim} N_r(\bar{\beta}; \Sigma)$.

Grazie alla riformulazione della funzione di verosimiglianza in termini di un modello gerarchico bayesiano, sarà possibile, usando la regola di Bayes, stimare le specifiche preferenze degli elettori e i parametri che descrivono la distribuzione mista N_r . La funzione di verosimiglianza sarà:

$$L(\beta_i, \bar{\beta}, \Sigma) = P(\text{dati} | \beta_i, \bar{\beta}, \Sigma) = \prod_{i=1}^N P(\text{dati} | \beta_i) \pi(\beta_i | \bar{\beta}, \Sigma) = \prod_{i=1}^N L_i(\beta_i) N(\beta_i | \bar{\beta}, \Sigma) \quad (4.20)$$

dove $L_i(\beta_i)$ è la verosimiglianza dei dati dell'i-esimo elettore condizionati sulle preferenze elettorali, mentre $N(\beta_i | \bar{\beta})$ è la distribuzione degli effetti casuali indicizzata dai parametri $\bar{\beta}$ e Σ .

Aver specificato la verosimiglianza è stato un passaggio dovuto per presentare la composizione della *posterior* dei parametri, ricavata della regola di Bayes al netto del fattore di normalizzazione a denominatore:

$$\pi(\beta_i, \bar{\beta}, \Sigma | \text{dati}) \propto \prod_{i=1}^N L_i(\beta_i) N(\beta_i | \bar{\beta}, \Sigma) \pi(\bar{\beta}, \Sigma) \quad (4.21)$$

Si nota che la quantità stimata, la posterior, è proporzionale al prodotto tra la funzione di verosimiglianza e la *prior*, ossia $\pi(\bar{\beta}, \Sigma)$. Dato che si sta seguendo un approccio bayesiano, è necessario specificare i *priors* sui parametri che descrivono la distribuzione delle preferenze optando per le distribuzioni a priori coniugate naturali, ovvero:

1. *Prior* di $\bar{\beta}$: $\bar{\beta} \sim N(0, V)$ per $V \rightarrow \infty$. Se $\beta_i \sim N_r(\bar{\beta}; \Sigma)$ è la distribuzione dei parametri nella popolazione, $\bar{\beta} \sim N(0, V)$ rappresenta le credenze a priori riguardanti la media di quella distribuzione. Esse sono descritte da una gaussiana non informativa in quanto, con V che tende a infinito, si avrà una normale quasi *flat* e ciò comporta che la *prior* eserciterà un'influenza minima sulla *posterior*;

2. *Prior* di Σ : $\Sigma \sim IW(\Psi, \nu)$. Le credenze a priori riguardanti la matrice di covarianza sono sintetizzate dai parametri Ψ e ν di una Wishart Inversa, ovvero di una distribuzione che rappresenta una generalizzazione multidimensionale di una Gamma¹⁷ inversa.

Abbiamo potuto usare queste particolari *prior* poiché il vettore delle preferenze è stato assunto distribuito secondo una Normale: con distribuzioni a priori di questo tipo, estrarre informazioni dalla *posterior* è molto più semplice [Solgaard and Hansen, 2003]. Per questo processo di estrazione di informazioni ci si avvale dell'algoritmo di Gibbs, in quanto permette di estrarre campioni relativi a un parametro condizionando su tutti gli altri parametri. Oltre ad approssimare la distribuzione congiunta di più variabili aleatorie, questo algoritmo è strutturato per approssimare le distribuzioni marginali delle diverse variabili casuali che compongono il modello.

Grazie alle proprietà e alla struttura dell'algoritmo di Gibbs, sarà dunque possibile stimare il vettore di parametri β_i , riuscendo così a risolvere la formula 4.19 per calcolare la probabilità che un certo elettore scelga un determinato partito.

¹⁷La distribuzione Gamma è una generalizzazione del χ^2 .

5. Applicazione del modello logistico multinomiale bayesiano alle preferenze elettorali dei cittadini europei

Questo capitolo ha lo scopo di riportare e commentare i risultati ottenuti dall'implementazione del modello logit multinomiale misto in chiave bayesiana nell'analisi della competizione politica a livello europeo, prendendo a riferimento l'VIII Legislatura del Parlamento europeo.

Verranno analizzati alcuni fattori che influenzano i *decision makers* delle principali democrazie avanzate europee nelle scelte di voto, prendendo come set di scelte i gruppi di partito interni al Parlamento europeo, chiamati anche famiglie europee. Questa scelta deriva dal fatto che, in questo modo, si superano le complicazioni derivanti dalle differenze nei diversi sistemi di partito nazionali e permette di raggruppare, in una famiglia comune, partiti di Nazioni diverse con orientamenti politici affini.

Tuttavia, all'interno delle singole democrazie, ci possono essere famiglie europee non rappresentate, in quanto vi è la possibilità che, a livello di singolo Paese, non sia stato votato nessun partito collegato a quella determinata famiglia o che, tra i partiti votati, non ve ne sia nessuno a essa associato.

Questo determina un'eterogeneità nel set di scelte ed è per questo motivo che il modello utilizzato è il MMLM in chiave bayesiana. Esso, come già spiegato nella sezione 4.3, è strutturato per rilassare l'assunzione IIA, permettendo di tenere in considerazione l'eterogeneità appena citata e restituendo stime più precise rispetto ai modelli logit multinomiali classici. Tali stime risultano più accurate in quanto consentono di superare alcuni limiti e al tempo stesso anche alcune violazioni del MMLM in chiave classica. Certamente un approccio non-parametrico consentirebbe di incrementare

anche l'accuratezza statistica delle stime ottenute.

Per interpretare correttamente modello, è utile inquadrare gli orientamenti politici dei vari gruppi di partito e fornire una breve descrizione delle variabili utilizzate.

Si procede, dunque, con una breve descrizione delle famiglie europee nella legislatura presa a riferimento¹:

- Il Partito Popolare Europeo (PPE) è il più vecchio gruppo politico europeo ed è, dal 1999, il gruppo di maggioranza relativa al parlamento europeo. Viene collocato nel centro destra e ha posizioni liberali, europeiste e si può definire, rimanendo in linea con la terminologia del primo capitolo di questa tesi, un “partito storico”;
- Il Gruppo dell’Alleanza Progressista dei Socialisti e dei Democratici al Parlamento europeo (S&D) è un gruppo più giovane, costituitosi nel 2009, con posizioni ideologiche di centro sinistra: progressismo, socialismo liberale e europeismo sono i tratti che lo caratterizzano;
- Il Gruppo dell’Alleanza dei Democratici e dei Liberali per l’Europa (ADLE) si colloca in centro e ha posizioni liberali ed europeiste;
- I Verdi Europei - Alleanza Libera Europea è un gruppo di centro sinistra che ha posizioni regionaliste, socialdemocratiche e ambientaliste;
- Il Gruppo dei Conservatori e dei Riformisti Europei (ECR) è collocato a destra ed è contrario ai processi di integrazione europea: euroscetticismo, nazionalismo e sovranismo sono i tratti che lo caratterizzano. Sono fortemente contrari all’integrazione europea anche il Gruppo Europa della Libertà e della Democrazia Diretta (EFDD) e il Gruppo Europa delle Nazioni e delle Libertà (ENF)², anche se quest’ultimo è considerato il più radicale in assoluto in termini di nazionalismo;
- Il Gruppo Confederale della Sinistra Unitaria Europea - Sinistra Verde Nordica (GUE/NGL), ha invece ideologie di estrema sinistra, è un gruppo di sinistra radicale.

¹I gruppi di partito possono variare da legislatura a legislatura, per questo è importante specificare la Legislatura presa in riferimento.

²Non è tra i gruppi presi in considerazione nell’analisi.

I dati utilizzati per l'implementazione del modello sono quelli relativi alla rilevazione dell'ESS (2018). Come già visto nel capitolo 2, l'ESS è un sondaggio di portata europea altamente professionale che sfrutta le migliori tecniche di campionamento per la raccolta dei dati ed è volto a indagare molteplici aspetti della vita dei cittadini europei: il *dataset* conta oltre 47000 osservazioni e 557 variabili.

Poiché in questo elaborato le variabili d'interesse sono quelle legate alla politica si è scelto di includere nel modello le seguenti:

- Variabile *“lrscle”*: è discreta e relativa a come si posizionerebbe un intervistato su una scala da 0 a 10, dove 0 sta per estrema sinistra e 10 per estrema destra;
- Variabile *“nwspol”*: è continua ed è relativa al tempo speso dall'intervistato a guardare, leggere e ascoltare notizie politiche e di attualità nell'arco di una giornata tipo;
- Variabile *“trstplt”*: è discreta ed è relativa alla fiducia che i cittadini dichiarano di avere nei politici. È misurata su una scala da 0 a 10, dove 0 indica sfiducia e 10 una fiducia totale;
- Variabile *“euftf”*: è discreta ed è molto interessante come misura di universalismo-particolarismo. Essa restituisce il posizionamento dell'intervistato rispetto all'integrazione europea su una scala da 0 a 10, dove 0 indica che l'integrazione europea ha già raggiunto livelli troppo elevati e 10 indica che il processo d'integrazione non è ancora sufficiente;
- Variabile *“imwbent”*: è una variabile discreta e, come la precedente, è interessante sotto la dimensione universalismo-particolarismo: viene chiesto all'intervistato se gli immigrati rendono lo Stato in cui vive un posto migliore o un posto peggiore. La scala è sempre da 0 a 10, con 0 che sta per “posto peggiore” e 10 per “posto migliore”;
- Variabile *“hinctnta”*: è discreta ed è relativa al reddito netto dei nuclei famigliari degli intervistati;

- Variabile “*eisced*”: è discreta ed è relativa al livello d’istruzione dell’intervistato, il quale è armonizzato secondo l’ISCED³;
- Variabile “*gndr*”: è discreta e restituisce il genere dell’intervistato;
- Variabile “*agea*”: restituisce l’età dell’intervistato.

L’analisi dei dati è stata effettuata utilizzando il software statistico R Studio (codice riportato nell’appendice C.1).

In primo luogo, ove possibile, a ogni partito che gli intervistati hanno dichiarato di aver votato alle ultime elezioni è stata associata la relativa famiglia europea (vedi appendice C.2), così da poter creare un vettore unico per il voto in termini di famiglie europee contenente il voto espresso in ogni Paese.

In seguito, con l’apposito comando, sono stati esclusi tutti gli intervistati con voto non palese e si sono perse circa metà delle 47086 osservazioni iniziali.

A questo punto, si è costruito un prospetto Excel per associare con lo stesso numero i Paesi con lo stesso set di scelte, ovvero con lo stesso numero e lo stesso tipo di partiti, per poter assegnare nei dati una variabile per poter considerare l’effetto aleatorio derivante dall’eterogeneità nel set di scelte.

Si è proceduto eliminando dal *dataset* anche i valori di voto residuali (per esempio “scheda bianca”), perdendo ulteriori 2114 osservazioni. Togliendo anche i valori mancanti nelle variabili utilizzate nel modello si sono perse ulteriori osservazioni, passando da 47086 osservazioni iniziali a 17571 finali.

Si è proceduto, infine, alla stima dei modelli e si riporta il confronto tra un modello a effetti fissi e lo stesso modello stimato introducendo gli effetti aleatori; quest’ultimo, per le caratteristiche analizzate alla fine del precedente capitolo, è stato implementato per catturare al meglio l’eterogeneità che caratterizza sia il set di alternative che i *decision-makers*.

Per il modello senza effetti aleatori è stata implementata una catena MCMC costruita attraverso l’algoritmo di Gibbs. Il *burn-in* è stato impostato a 1000, scartando quindi questo numero di campioni, in quanto relativi al periodo antecedente al raggiungi-

³È uno standard creato dall’UNESCO come sistema internazionale di classificazione dei livelli di istruzione.

mento della distribuzione stazionaria e quindi poco rappresentativi della distribuzione obiettivo. Il numero di iterazioni della catena è stato impostato a 11000.

Per il secondo modello, il *burn-in* è stato impostato a 5000, il numero di iterazioni a 45000 e il “*thin*”, ovvero un parametro che specifica ogni quanto viene selezionata un’osservazione, utile per limitare l’autocorrelazione e favorire la convergenza della catena, è stato impostato a 5. Le figure 5.1 e 5.2, contengono rispettivamente le stime ottenute per il modello a effetti fissi e a effetti aleatori. Si procede con il commento dei risultati ottenuti.

I primi otto coefficienti stimati, le medie a posteriori delle distribuzioni ottenute dai metodi iterativi di campionamento, rappresentano le intercette relative ai gruppi di partito comparati con il gruppo di riferimento ADLE e indicano il log dell’ODDS, ovvero il logaritmo della probabilità associata al votare quel determinato partito rispetto alla probabilità di votare il partito di riferimento. Il segno del coefficiente indica se la probabilità della categoria presa in considerazione aumenta o diminuisce rispetto alla probabilità della categoria di riferimento e per identificare l’intensità dell’effetto generato dalla variabile è necessario calcolare l’ODDS RATIO, il che equivale a calcolare l’ $\exp(\text{ODDS})$, ovvero l’esponenziale dei coefficienti.

Si può notare che il primo modello a intercette fisse differisce dal secondo modello a intercette aleatorie in termini di significatività: nel primo praticamente tutti i coefficienti delle intercette sono statisticamente significativi, con una significatività leggermente più debole per i Verdi, mentre nel secondo modello a intercette aleatorie sono significativi solo i coefficienti associati al PPE e all’S&D. Questo aspetto significa che l’effetto aleatorio introdotto nel secondo modello influenza fortemente il modello e che, quindi, l’effetto casuale derivante dall’eterogeneità nel set di scelte è rilevante. In particolare, l’introduzione di un effetto aleatorio introduce naturalmente maggiore variabilità e quanto emerge è quindi naturale.

Inserire gli effetti aleatori equivale, infatti, ad aggiungere ulteriori elementi che hanno una propria variabilità: è dunque ragionevole aspettarsi una riduzione della significatività. È importante tenere conto di questo aspetto per non cadere nell’errore di assumere come significativi degli elementi che, in realtà, non lo sono. Ciò è anche in linea con la realtà, in quanto ulteriori elementi di incertezza derivanti da set di scelta non omogenei rendono ancora più complicato individuare gli elementi che effettivamente hanno

un'influenza sul fenomeno d'interesse. Non tenere conto dell'eterogeneità derivante dal fatto che in Paesi diversi siano rappresentate famiglie europee diverse potrebbe indurre a considerare come significativi degli elementi che in realtà non lo sono.

Si può notare che anche la significatività relativa alle altre variabili differisce in maniera sostanziale tra i due modelli ed è sempre da interpretare come il log della probabilità di quella determinata categoria rispetto alla categoria di riferimento.

La categoria di riferimento è in genere il primo livello della variabile presa in considerazione. Si riportano solamente le categorie di riferimento associate alle variabili rilevanti del secondo modello:

- per l'istruzione la categoria “non è stato possibile armonizzare il livello d'istruzione secondo i parametri ISCED”;
- per il genere è “maschi”;
- Per il reddito è il primo decile;
- Per il posizionamento politico è la sinistra;
- Per la variabile relativa all'immigrazione è “gli immigrati rendono il Paese un posto peggiore”;
- Per la variabile relativa alla fiducia nei politici è “completa sfiducia nei partiti”.

Nel secondo modello, le variabili che influenzano il voto in modo significativo sono molte meno, a dimostrazione del fatto che l'introduzione di ulteriore incertezza comporta l'esclusione di alcune variabili che il modello precedente assumeva come significative.

Modello 4_no-effetti-aleatori	Media a posteriori	l-95% C.I	u-95% C.I.	eff.samp	pMCMC
ECR	1.225e-01	5.282e-02	1.928e-01	2.639	<0,001***
EFDD	1.505e-01	7.817e-02	2.181e-01	2.677	<0,001***
ENL	1.211e-01	4.474e-02	1.887e-01	2.599	<0,001***
NGL	1.092e-01	3.785e-02	1.790e-01	2.611	<0,001***
PPE	1.291e-01	6.585e-02	1.919e-01	2.901	<0,001***
S&D	1.139e-01	5.189e-02	1.795e-01	2.869	<0,001***
VERDI-ALE	9.767e-02	2.468e-02	1.674e-01	2.668	0,002**
Altro	1.201e-01	4.715e-02	1.903e-01	2.560	<0,001***
Età	-3.382e-04	-6.032e-04	-4.489e-05	2.433	<0,001***
Genere_femmine	5.283e-02	4.023e-02	6.069e-02	3.007	<0,001***
Istruzione_ES-ISCED I	-1.207e-01	-1.701e-01	5.889e-02	3.965	<0,001***
Istruzione_ES-ISCED II	-2.087e-01	-2.516e-01	1.525e-01	4.833	<0,001***
Istruzione_ES-ISCED IIIb	-3.419e-01	-4.071e-01	-2.778e-01	2.716	<0,001***
Istruzione_ES-ISCED IIIa	-1.301e-01	-1.972e-01	-6.598e-02	3.311	<0,001***
Istruzione_ES-ISCED IV	-2.730e-01	-3.317e-01	-2.236e-01	4.167	<0,001***
Istruzione_ES-ISCED V1	-2.393e-01	-2.908e-01	-1.769e-01	3.619	<0,001***
Istruzione_ES-ISCED V2	-3.040e-01	-3.605e-01	-2.525e-01	4.028	<0,001***
Reddito_2°decile	1.350e-02	-2.160e-02	3.644e-02	2.120	0,466
Reddito_3°decile	-1.938e-03	-2.666e-02	1.499e-02	4.043	0,958
Reddito_4°decile	1.220e-02	-2.390e-02	3.519e-02	2.283	0,414
Reddito_5°decile	1.412e-02	-4.875e-03	3.558e-02	3.716	0,344
Reddito_6°decile	2.843e-02	-4.722e-04	4.664e-02	1.871	0,05
Reddito_7°decile	-1.615e-02	4.551e-02	5.997e-03	2.045	0,228
Reddito_8°decile	-9.770e-03	-3.235e-02	7.433e-03	2.762	0,588
Reddito_9°decile	1.891e-02	-3.176e-03	3.771e-02	2.764	0,204
Reddito_10°decile	9.575e-02	5.967e-02	1.167e-01	2.335	<0,001***

Posizionamento Politico centrosinistra	-5.195e-03	-1.280e-02	1.490e-02	7.337	0,244
Posizionamento Politico centro	-7.739e-02	9.926e-02	-4.102e-02	2.572	<0,001***
Posizionamento Politico centrodestra	3.022e-03	-1.690e-02	2.677e-02	1.997	0,862
Posizionamento Politico destra	5.067e-02	2.738e-02	8.249e-02	2.096	<0,001***
Interesse politico	7.236e-05	5.329e-05	9.033e-05	6.667	<0,001***
Fiducia nei politici	-6.286e-03	-9.182e-03	-3.232e-03	1.511	<0,001***
Volontà Integrazione Europea	1.863e-02	1.483e-02	2.270e-02	1.261	<0,001***
Pro-immigrazione_1	2.207e-02	-3.075e-02	8.447e-02	1.623	0,648
Pro-immigrazione_2	4.118e-02	2.131e-02	7.684e-02	3.072	<0,001***
Pro-immigrazione_3	3.922e-02	1.223e-02	5.555e-02	2.334	<0,001***
Pro-immigrazione_4	1.276e-03	-1.404e-02	2.705e-02	3.349	0,922
Pro-immigrazione_5	-3.953e-03	-2.400e-02	1.320e-02	1.743	0,796
Pro-immigrazione_6	-1.944e-02	-5.572e-02	7.957e-03	0.890	0,384
Pro-immigrazione_7	5.235e-02	3.433e-02	7.272e-02	2.982	<0,001***
Pro-immigrazione_8	2.604e-02	1.907e-04	4.853e-02	1.670	0,030*
Pro-immigrazione_9	6.843e-02	5.083e-02	8.692e-02	4.138	<0,001***

Figura 5.1: Modello a effetti fissi.

Modello 4_con- effetti-aleatori	Media a posteriori	l-95% C.I	u-95% C.I.	eff.samp	pMCMC
ECR	0.5704187	-1.9042057	3.1208758	422.28	0.64350
EFDD	-1.2974641	3.5943882	1.0420478	126.27	0.27350
ENL	-1.3781049	-4.2843174	1.1762367	70.75	0.33425
NGL	1.1776389	-0.7589810	3.2418785	300.09	0.24400
PPE	4.1475451	2.4163561	5.9000868	246.54	< 1e-04***
S&D	3.8400366	1.9191547	5.8475371	330.61	0.00025***
VERDI-ALE	1.1006674	-0.7552684	2.7907664	252.43	0.22450
Altro	0.5824281	-1.2729469	2.4619300	270.58	0.53125
Età	-0.0039033	-0.0080980	-0.0003773	107.91	0.04650 *
Genere_femmine	0.0393437	-0.0542908	0.1364001	181.06	0.42650
Istruzione_ES-ISCED I	-0.4346053	-1.4970146	0.5714781	126.18	0.42525
Istruzione_ES-ISCED II	-0.5034641	-1.6159506	0.4999139	119.48	0.34700
Istruzione_ES-ISCED IIIb	-0.1659103	-1.1920822	0.8785718	124.56	0.76375
Istruzione_ES-ISCED IIIa	-0.3869110	-1.4225814	0.6416190	123.45	0.47125
Istruzione_ES-ISCED IV	-0.5268788	-1.5777500	0.5145738	121.43	0.32250
Istruzione_ES-ISCED V1	-0.4806140	-1.5920675	0.5303340	119.07	0.37800
Istruzione_ES-ISCED V2	-0.6048114	-1.7290372	0.4065875	119.55	0.25350
Reddito_2°decile	0.2194366	-0.0487631	0.5027104	116.81	0.11775
Reddito_3°decile	0.1654543	-0.0782444	0.4208493	153.62	0.18650
Reddito_4°decile	0.2212813	-0.0304188	0.4698312	155.70	0.08850
Reddito_5°decile	0.2630087	-0.0060751	0.5220005	145.08	0.06700
Reddito_6°decile	0.3170929	0.0615667	0.5732160	157.12	0.01750 *
Reddito_7°decile	0.2275668	-0.0285302	0.4711477	174.58	0.07025
Reddito_8°decile	0.2488510	-0.0396096	0.5366260	120.57	0.08775
Reddito_9°decile	-0.0603633	-0.3284502	0.2202696	141.65	0.65850
Reddito_10°decile	-0.2206408	-0.4860214	0.0329957	156.12	0.09825
Posizionamento Politico_centrosinistra	-0.7787968	-0.9805879	-0.5761390	106.10	< 1e-04***
Posizionamento Politico_centro	-1.4217817	-1.6205343	-1.2080114	108.86	< 1e-04***
Posizionamento Politico_centrodestra	-1.5233833	-1.7501989	-1.3161010	105.33	< 1e-04***
Posizionamento Politico_destra	-0.9469211	-1.2218802	-0.7105429	82.08	< 1e-04***

Interesse politico	0.0008785	0.0004586	0.0013405	113.07	< 1e-04***
Fiducia nei politici	-0.1037214	-0.1286749	-0.0747467	102.78	< 1e-04***
Volontà Integrazione Europea	-0.0139798	-0.0581104	0.0358952	96.69	0.55350
Pro-immigrazione_1	-0.1276411	-0.5996960	0.3129116	107.21	0.58825
Pro-immigrazione_2	-0.2050191	-0.5909817	0.1320083	113.17	0.24800
Pro-immigrazione_3	0.0385096	-0.2974525	0.3842298	95.29	0.81875
Pro-immigrazione_4	-0.2560408	-0.5627861	0.0402842	109.88	0.08400
Pro-immigrazione_5	-0.4248212	-0.6952278	-0.1846214	110.68	0.00025***
Pro-immigrazione_6	-0.3593351	-0.6328284	-0.1021482	140.11	0.00725 **
Pro-immigrazione_7	-0.1647375	-0.4417783	0.0999645	118.57	0.22975
Pro-immigrazione_8	-0.2004513	-0.4788348	0.0604037	157.08	0.13400
Pro-immigrazione_9	-0.1800745	-0.5350302	0.1721082	141.59	0.31350

Figura 5.2: Modello a effetti aleatori.

6. Conclusioni

All'inizio di questo elaborato si è analizzato come le trasformazioni socio-economiche nelle principali democrazie europee abbiano influito sulle dinamiche di voto. Si è visto che i processi di globalizzazione e la terziarizzazione del mercato del lavoro hanno spostato la battaglia politica su nuovi fronti rispetto ai classici temi legati alla religione e alla classe sociale. Infatti, seguendo le logiche della teoria dei *cleavages*, sono state individuate due dimensioni dello spazio politico: una dimensione particolarismo-universalismo, la quale contrappone posizioni degli elettori su temi legati agli effetti dei processi di globalizzazione, e una dimensione Stato-mercato che contrappone gli elettori a favore dell'intervento dello Stato nell'economia, inteso principalmente in termini redistributivi, e gli elettori con posizioni più liberiste.

Si è visto come, per supportare queste posizioni a livello politico, sono nati nuovi partiti che, nel giro di pochi anni, hanno assunto un peso politico rilevante all'interno delle istituzioni europee e hanno messo in crisi i partiti tradizionali. Questi ultimi, seppur in maniera limitata, hanno dovuto dare importanza ai nuovi temi legati alle due dimensioni precedentemente descritte per cercare di non perdere una parte importante di elettorato.

In questa tesi è stata proposta una nuova metodologia per analizzare il voto europeo che, tramite l'introduzione degli effetti aleatori, permette di gestire i problemi derivanti dal fatto che non tutte le famiglie europee sono rappresentate nei vari Paesi europei.

Questo aspetto determina eterogeneità nel set di scelta degli elettori, caratteristica che rende inopportuni modelli che pesano gli attributi delle possibili scelte allo stesso modo. Si pensi a un Paese 1 in cui vi siano solo due partiti (ognuno dei quali associati ad un diverso gruppo di partito europeo) e a un Paese 2 in cui vi siano sette partiti appartenenti a sette famiglie europee, tra cui rientrano i due gruppi rappresentati nel Paese 1: gli elettori dei due Paesi difficilmente peseranno gli attributi dei vari partiti allo stesso modo, in quanto nel primo vi è molta meno scelta. Inoltre, il modello

proposto tiene conto del fatto che esistono fattori non osservabili o non osservati che influenzano le decisioni di voto degli elettori e che, dunque, elettori con caratteristiche simili potrebbero avere posizioni politiche fortemente divergenti.

Tramite la comparazione delle stime risultanti dai due modelli implementati nel capitolo precedente, i quali differiscono solamente per l'introduzione degli effetti aleatori, si è evidenziato come l'eterogeneità nel set di scelta giochi un ruolo rilevante nelle dinamiche di voto europee. I due modelli differiscono sostanzialmente in termini di significatività sia delle intercette che delle altre variabili e ciò sta a significare che l'ulteriore componente di varianza introdotta nel secondo modello gioca un ruolo rilevante: essa permette di stimare un modello più parsimonioso e aderente alla realtà.

Il modello che permette di effettuare analisi di questo tipo è strutturato secondo le logiche inferenziali bayesiane e, per questo motivo, è stato sviluppato un capitolo nel quale se ne presentano le principali peculiarità e si fornisce una panoramica generale sui metodi di supporto necessari per implementarle.

Essendo, inoltre, un modello non lineare, esso è stato introdotto gradualmente per permettere di capirne le differenze con i modelli lineari, spiegando inizialmente le caratteristiche della versione più semplice, ovvero del modello *logit*, per arrivare alla versione più complessa proposta in questa tesi.

La parte metodologica è stata pensata per coloro che, per la prima volta, si avvicinano a questi temi statistici e, per questo motivo, sono state strutturate delle note a supporto che potrebbero aiutare la comprensione dei temi trattati.

Un possibile sviluppo del presente lavoro risiede nella possibilità di specificare invece che un modello parametrico, un modello Bayesiano semi-parametrico o non-parametrico al fine di recuperare maggiore flessibilità e, come dimostrato in diverse applicazioni empiriche presenti nella letteratura di riferimento, maggiore significatività delle stime rispetto al modello parametrico.

Bibliografia

- [Agnoli, 2002] Agnoli, P. (2002). Sulla probabilità.
- [Andrieu and Thoms, 2008] Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive mcmc. *Statistics and computing*, 18(4):343–373.
- [Anthony et al., 2011] Anthony, R. N., Hawkins, D. F., Macri, D. M., Merchant, K. A., et al. (2011). *Sistemi di controllo. Analisi economiche per le decisioni aziendali*. McGraw-Hill.
- [Azzalini, 2004] Azzalini, A. (2004). *Inferenza statistica: Una presentazione basata sul concetto di verosimiglianza*. Springer Science & Business Media.
- [Baldi, 2003] Baldi, P. (2003). *Introduzione alla probabilità con elementi di statistica*. McGraw-Hill.
- [Ballone and Colagrande, 1994] Ballone, E. and Colagrande, V. (1994). La dimensione campionaria nella verifica di ipotesi statistiche. *Ratio Mathematica*, 7(1):25–38.
- [Beramendi et al., 2013] Beramendi, P., Häusermann, S., Kitschelt, H., and Kriesi, H. (2013). The politics of advanced capitalism: An introduction.
- [Bertsekas and Tsitsiklis, 2008] Bertsekas, D. and Tsitsiklis, J. (2008). Introduction to probability, ser. *Athena Scientific optimization and computation series*. Athena Scientific.
- [Blanchard et al., 2014] Blanchard, O., Amighini, A., and Giavazzi, F. (2014). *Macroeconomia: una prospettiva europea*. Bologna: Il Mulino.
- [Capano et al., 2014] Capano, G., Piattoni, S., Raniolo, F., and Verzichelli, L. (2014). *Manuale di scienza politica, il mulino*.

- [Capello, 2004] Capello, R. (2004). *Economia regionale*. Il mulino.
- [Caravenna and Dai Pra, 2013] Caravenna, F. and Dai Pra, P. (2013). *Probabilità: Un'introduzione attraverso modelli e applicazioni*, volume 67. Springer Science & Business Media.
- [Carlo, 2004] Carlo, C. M. (2004). Markov chain monte carlo and gibbs sampling. *Lecture notes for EEB*, 581.
- [Corbetta et al., 2001] Corbetta, P., Gasperoni, G., and Pisati, M. (2001). *Statistica per la ricerca sociale*. Il mulino.
- [Crisma, 2006] Crisma, L. (2006). *Probabilità coerenti*. EUT Edizioni Università di Trieste.
- [Cucciniello et al., 2020] Cucciniello, M., Fattore, G., Longo, F., Ricciuti, E., and Turrini, A. (2020). *Management pubblico*. EGEA spa.
- [Dancygier and Walter, 2015] Dancygier, R. M. and Walter, S. (2015). Globalization, labor market risks, and class cleavages.
- [De Bettin, 2014] De Bettin, A. (2014). Validità dei test basati sulla verosimiglianza penalizzata in modelli log-lineari.
- [De Finetti, 1931] De Finetti, B. (1931). Sul significato soggettivo della probabilità. *Fundamenta mathematicae*, 17(1):298–329.
- [De Grassi et al., 2009] De Grassi, M., Naticchia, B., Giretti, A., and Carbonari, A. (2009). *Reti bayesiane con applicazioni all'edilizia e alla gestione del territorio*. FrancoAngeli.
- [Durio and Isaia, 2009] Durio, A. and Isaia, E. (2009). *Calcolo delle Probabilità e Variabili Casuali*. Dipartimento di Statistica e Matematica Applicata “Diego de Castro”.
- [Epifani, 2009] Epifani, I. (2009). Appunti delle lezioni del corso di statistica (2l) per gli allievi inf e tel, aa 2008/2009 teoria della stima puntuale.

- [Erto et al., 2012] Erto, P., Giorgio, M., and Iervolino, I. (2012). Probabilità e rischio. *Ambiente, rischio, comunicazione-Decidere nell'incertezza*, 4:64.
- [Fiebig et al., 2010] Fiebig, D. G., Keane, M. P., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science*, 29(3):393–421.
- [Friedman et al., 1997] Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine learning*, 29(2-3):131–163.
- [Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.
- [Grossman and Domingos, 2004] Grossman, D. and Domingos, P. (2004). Learning bayesian network classifiers by maximizing conditional likelihood. In *Proceedings of the twenty-first international conference on Machine learning*, page 46.
- [Guttag, 2016] Guttag, J. (2016). *Introduction to computation and programming using Python: With application to understanding data*. MIT Press.
- [Häusermann and Kriesi, 2015] Häusermann, S. and Kriesi, H. (2015). What do voters want? dimensions and configurations in individual-level preferences and party choice. *The politics of advanced capitalism*, pages 202–230.
- [Hensher and Greene, 2003] Hensher, D. A. and Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, 30(2):133–176.
- [Hooghe and Marks, 2018] Hooghe, L. and Marks, G. (2018). Cleavage theory meets europe’s crises: Lipset, rokkan, and the transnational cleavage. *Journal of European Public Policy*, 25(1):109–135.
- [Iannazzo, 2007] Iannazzo, S. (2007). Le tecniche statistiche bayesiane e la loro applicazione in modelli di simulazione probabilistica. *Farmeconomia e percorsi terapeutici*, 8(1):5–13.
- [James et al., 2013] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

- [Kolomogoroff, 2013] Kolomogoroff, A. (2013). *Grundbegriffe der wahrscheinlichkeitstheorie*, volume 2. Springer-Verlag.
- [Laplace, 1998] Laplace, P.-S. (1998). *Pierre-Simon Laplace Philosophical Essay on Probabilities: Translated from the fifth French edition of 1825 With Notes by the Translator*, volume 13. Springer Science & Business Media.
- [Lesaffre and Lawson, 2012] Lesaffre, E. and Lawson, A. B. (2012). *Bayesian biostatistics*. John Wiley & Sons.
- [Locatelli et al., 2003] Locatelli, I., Lichtenstein, P., Yashin, A. I., et al. (2003). *A Bayesian correlated frailty model applied to Swedish breast cancer data*. Università commerciale Luigi Bocconi.
- [Nava et al., 2016] Nava, C. R., Carota, C., and Colombino, U. (2016). Bayesian methods for microsimulation models. In *International Conference on Bayesian Statistics in Action*, pages 193–202. Springer.
- [Neal, 1993] Neal, R. M. (1993). *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto Toronto, Ontario, Canada.
- [Pastore, 2013] Pastore, M. (2013). Le tecniche di simulazione monte carlo. *Barbaranelli, s. ingoglia (a cura di), I modelli di equazioni strutturali: temi e prospettive. Milano: Led*, pages 295–324.
- [Piccinato, 2009] Piccinato, L. (2009). *Metodi per le decisioni statistiche*. Springer Science & Business Media.
- [Prügel-Bennett, 2020] Prügel-Bennett, A. (2020). *The Probability Companion for Engineering and Computer Science*. Cambridge University Press.
- [Roberts et al., 2001] Roberts, G. O., Rosenthal, J. S., et al. (2001). Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, 16(4):351–367.
- [Rossi et al., 2012] Rossi, P. E., Allenby, G. M., and McCulloch, R. (2012). *Bayesian statistics and marketing*. John Wiley & Sons.

- [Sanfilippo, 2015] Sanfilippo, G. (2015). L'incertezza insegna.
- [Shiryaev, 2019] Shiryaev, A. N. (2019). Markov chains. In *Probability-2*, pages 237–312. Springer.
- [Solgaard and Hansen, 2003] Solgaard, H. S. and Hansen, T. (2003). A hierarchical bayes model of choice between supermarket formats. *Journal of retailing and Consumer Services*, 10(3):169–180.
- [Sottile, 2015] Sottile, E. (2015). Costruzione di modelli di scelta discreta ibridi per misurare gli effetti delle strategie informative sulla scelta del modo di viaggio contestualmente agli aspetti latenti.
- [Treccani, 1986] Treccani, G. (1986). Vocabolario della lingua italiana. *Istituto della Enciclopedia Italiana*.
- [Tsitsiklis, 2010] Tsitsiklis, J. (2010). *6.041 Probabilistic Systems Analysis and Applied Probability*. Massachusetts Institute of Technology: MIT OpenCourseWare.
- [Yamamoto, 2011] Yamamoto, T. (2011). A multinomial response model for varying choice sets, with application to partially contested multiparty elections. Technical report, Working paper.

A. Appendice

A.1 Fondamenti della probabilità

La teoria delle probabilità è quella branca della matematica che, modellizzando *esperimenti casuali*, si occupa di studiare le frequenze con cui si verificano gli *eventi*.

Per *esperimenti casuali* si intendono quegli esperimenti che possono essere ripetuti indefinitivamente sotto le stesse condizioni e ai quali è possibile esprimere, in modo esaustivo e prima di eseguirli, i loro possibili esiti, ma senza poter conoscere con certezza quali si realizzeranno [Durio and Isaia, 2009].

L'insieme di tutti i possibili risultati di un esperimento casuale, di cui non conosciamo l'esito, sarà indicato con Ω , mentre ω_α rappresenterà i singoli risultati, o *eventi elementari*. Per *evento* ci riferiremo invece ad un sottoinsieme proprio o improprio¹ dello *spazio degli eventi* [Prügel-Bennett, 2020].

Ad esempio, scommettendo in una schedina l'esito di una partita di calcio, l'insieme degli eventi sarà $\Omega = \{1, X, 2\}$, dove 1 è associato alla vittoria della squadra sulla quale abbiamo scommesso, X al pareggio e 2 alla sconfitta. L'evento, d'altro canto, potrebbe essere $\omega = \{1, X\}$, ovvero un sottoinsieme di Ω associato alla vittoria e al pareggio della squadra.

Per fornire un metodo di calcolo della probabilità per gli eventi, ci baseremo sull'impostazione assiomatica di Kolmogorov, il quale, nella sua Opera ², agli inizi degli anni Trenta del secolo scorso, ne sviluppa un approccio innovativo e in termini puramente matematici. Successivamente, affronteremo anche i diversi approcci alla defini-

¹Il sottoinsieme proprio contiene solo una parte degli eventi elementari mentre il sottoinsieme improprio può essere solamente vuoto o coincidere con Ω .

²L'opera è intitolata "Grundbegriffe der wahrscheinlichkeitsrechnung", per maggiori approfondimenti [Kolmogoroff, 2013].

zione di probabilità, per soffermarci poi su una in particolare, chiamata *soggettivista* [Piccinato, 2009].

A.1.1 Impostazione assiomatica

Probabilità come funzione d'insieme. La probabilità è considerata come una funzione d'insieme avente come dominio una σ -algebra associata ad Ω , chiamato *spazio probabilizzabile*, ed operante in conformità di tre assiomi, chiamati appunto *assiomi di Kolmogorov*.

Quando si vuole definire uno *spazio probabilizzabile*, è infatti opportuno selezionare tra tutti gli eventi solamente quelli sottostanti a determinate proprietà di chiusura rispetto ad alcune operazioni insiemistiche, in particolare quelle di unione e complemento, tali da costituire una σ -algebra, \mathcal{A} . [Piccinato, 2009] [Durio and Isaia, 2009].

\mathcal{A} , o *dominio di probabilità*, è dunque la collezione di tutti i possibili sottoinsiemi dello spazio degli eventi Ω se:

- $\Omega \in \mathcal{A}$;
- $\forall A \in \Omega \rightarrow A \in \mathcal{A}$ e $A^c \in \mathcal{A}$, dove A^c è il complementare ed è dato da $\Omega \setminus \{A\}$;
- $\bigcup_{i=1}^n A_i \in \mathcal{A}$.

Spazio probabilizzabile, Queste tre proprietà [Baldi, 2003], dove A indica un insieme generico, ci permettono di associare ad Ω un insieme di suoi sottoinsiemi che racchiuda in sé tutti i possibili modi con cui si vorrà interpretare l'esito dell'esperimento, così da poter definire anche la composizione dello *spazio probabilizzabile*, ovvero (Ω, \mathcal{A}) .

Ricapitolando, dato un *esperimento aleatorio*, una volta individuato l'*insieme dei possibili esiti* ed averci associato un'*algebra*, è possibile costruire lo *spazio probabilizzabile* sul quale sarà possibile definire la *probabilità* [Durio and Isaia, 2009].

Dato (Ω, \mathcal{A}) , si definisce *misura della probabilità* una qualsiasi funzione: $P: \mathcal{A} \rightarrow [0, 1]$, che soddisfi tali assiomi³:

³Assiomi di Kolmogorov.

- $\forall A \in \mathcal{A} \rightarrow P(A) \geq 0$; *assioma della non negatività*.
- $\forall A_1 \dots A_n$ tale che $A_i \cap A_j = \emptyset$ per $i \neq j$, allora $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, nel caso in cui $n = \infty$ si avrà $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$; *assioma dell'unione*.
- Per $P(A_2) > 0 \rightarrow P(A_1|A_2) = \frac{A_1 \cap A_2}{A_2}$, da cui si ricava: $P(A_1 \cap A_2) = P(A_2) \cdot P(A_1|A_2) = P(A_2 \cap A_1) = P(A_1) \cdot P(A_2|A_1)$, ovvero il *principio delle probabilità composte*⁴.
- $P(\Omega) = 1$; *assioma della certezza*.

Da questi assiomi sono deducibili numerose proprietà [Kolomogoroff, 2013], tra cui le seguenti:

1. $P(\emptyset) = 0$
2. $\forall A \in \mathcal{A} \rightarrow P(A) \leq 1$
3. $A_1 \subseteq A_2 \rightarrow P(A_1) \leq P(A_2)$
4. $A \cup A^c = \Omega, A \cap A^c = \emptyset, \forall A \in \mathcal{A} \rightarrow P(A^c) = 1 - P(A)$
5. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
6. $P(A_1 \cap A_2) \leq \min(P(A_1), P(A_2))$

È inoltre opportuno segnalare che nel caso l'algebra sia definita su \mathbb{R} , essa viene definita *σ -Borel algebra* e lo spazio probabilizzabile diventa (\mathbb{R}, β) . In questo caso la funzione avrà per dominio β e per codominio l'insieme dei numeri reali.

Spazio probabilizzato, Ora il problema si cela nell'individuazione della misura di probabilità P che si intende adottare in modo da definire univocamente la terna ordinata (Ω, \mathcal{A}, P) , detta anche *spazio probabilizzato*; in altri termini, bisogna scegliere quale funzione di probabilità utilizzare per misurare gli eventi che compongono l'algebra.

Nel caso in cui l'insieme dei possibili esiti dell'esperimento aleatorio sia *finito*, ovvero quando vi sia la possibilità di conoscere il numero totale di eventi elementari⁵, e quando

⁴Se i due eventi sono stocasticamente indipendenti, ovvero il verificarsi di uno non modifica la probabilità che si verifichi l'altro, si avrà che $P(A_1|A_2) = P(A_1)$ e $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$.

⁵Si ricordi che per eventi elementari si intendono i singoli risultati dell'esperimento.

l'algebra associata coincida con il suo insieme delle parti, si attribuisce la probabilità degli eventi elementari in conformità agli assiomi. Successivamente si potrà attribuire la probabilità ad un qualsiasi evento dell'algebra considerandolo come unione di eventi elementari; sommando dunque le probabilità associate agli eventi elementari, sarà possibile ricavare la probabilità dell'evento preso in considerazione [Durio and Isaia, 2009]. Se nella terna (Ω, \mathcal{A}, P) la misura di probabilità assegna uguale probabilità ad ogni evento elementare, si parlerà di spazio probabilizzato equiprobabile. Questa specifica fattispecie, sarà il perno sul quale si articola l'approccio classico alla definizione stessa di probabilità. Vedremo che essa si determinerà come rapporto tra il numero di casi favorevoli e il numero di casi possibili [Baldi, 2003] [Durio and Isaia, 2009].

Anche nel caso di un Ω con cardinalità *infinita numerabile* la costruzione dello spazio probabilizzato seguirà lo stesso schema. Quando invece l'insieme dei possibili esiti coincide con l'insieme dei numeri reali o con un suo sottoinsieme, dunque presenti cardinalità *infinita non numerabile*, l'algebra contenente gli eventi a cui associare una probabilità è quella di Borel; tale associazione risulterà più complessa rispetto ai casi precedenti anche se il meccanismo è praticamente analogo. Si assegnerà la probabilità alla famiglia di intervalli che genera β per poi estenderla a tutti i suoi elementi, [Baldi, 2003] [Durio and Isaia, 2009]. Nei primi due casi, avremo uno spazio degli eventi *discreto* mentre nel caso di cardinalità infinita, lo spazio degli eventi sarà *continuo*.

Con l'impostazione assiomatica, Kolmogorov si dissociò dalle discussioni storiche legate concetto di probabilità, per concentrare i suoi sforzi nel formulare una teoria comune a tutte le scuole di pensiero. Le varie concezioni di probabilità, partendo dall'impostazione assiomatica, ne arricchiranno il contenuto e determineranno alcuni casi particolari, motivo per cui nella prossima sezione verranno trattate. In particolare, l'approccio soggettivista è alla base dei modelli che discussi e applicati in questo elaborato.

A.1.2 Concetto di probabilità: vari approcci

Essendo il *calcolo delle probabilità* una scienza matematica, è stato necessario elencare i postulati, le regole di coerenza che lo governano e, successivamente, le varie proprietà, in modo da legittimare matematicamente i criteri utilizzati da colui che applica questa scienza.

Un calcolo probabilistico che segua tali regole restituisce sicuramente un risultato scientifico inconfutabile.

Tuttavia, per darne validità pratica, è fondamentale che vi sia una forte professionalità e completa imparzialità da parte di colui che effettua il calcolo e che le ipotesi adottate vengano sempre esplicitate. L'analista deve limitarsi a *interpretare* ciò che i dati comunicano, evitando di *manipolarli* per ottenere un risultato sperato [Erto et al., 2012].

Proprio nell'interpretare gli *output* del calcolo probabilistico, che ricordiamo avere le radici nella teoria assiomatica, tornerà utile approfondire le varie definizioni di probabilità, in quanto ognuna di esse porterà ad una differente interpretazione dei risultati ottenuti.

A.1.2.1 Approccio classico

La *definizione classica* è quella formalizzata dal matematico Pierre Simon Laplace nel 1779 e fa riferimento ad un esperimento che può avere N esiti diversi, equiprobabili e incompatibili tra di loro; il verificarsi di uno non implica che gli altri non possano verificarsi [Laplace, 1998].

Associando a N_A il numero di esiti che fanno verificare l'evento A e a N la cardinalità di Ω , ovvero il numero associato all'insieme di tutti gli esiti possibili, la probabilità che si verifichi A è definita mediante tale equazione:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{N_A}{N} \quad (\text{A.1})$$

Un esempio può essere utile per concretizzare tale definizione. Assumendo il lancio di una moneta, i possibili esiti di tale lancio sono testa o croce e sono equiprobabili. In questo caso $|\Omega|$, ovvero la cardinalità di tutti i possibili esiti, è pari a due. Se vogliamo ricavare la probabilità di ottenere testa, l'equazione (3.1) diventa:

$$P(T) = \frac{N_T}{|\Omega|} = \frac{1}{2} \quad (\text{A.2})$$

Un ulteriore esempio è rappresentato dal lancio di un dado. Se non è truccato, la probabilità per ogni faccia sarà la medesima. Si vuole identificare la probabilità che esca un numero pari dal lancio; in questo caso $|\Omega| = \{1, 2, 3, 4, 5, 6\}$ e la probabilità che

si concretizzi l'evento desiderato sarà pari a:

$$P(2, 4, 6) = \frac{3}{6} = \frac{1}{2} \quad (\text{A.3})$$

Come detto in precedenza, l'approccio di Laplace nell'esprimere il concetto di probabilità non è l'unico; le prossime sezioni saranno infatti focalizzate su altre definizioni di probabilità, più attuali e più esaurienti, in quanto formulate per superare alcuni limiti della definizione classica. In particolare, si enunceranno i concetti relativi alla *probabilità frequentista* e alla *probabilità soggettivista/bayesiana*.

A.1.2.2 Approccio frequentista

La definizione classica, pur essendo una definizione di facile applicazione e funzionante in determinati ambiti, presenta alcune debolezze.

Dal punto di vista strettamente formale, essa è una definizione circolare in quanto, volendo definire cos'è la probabilità, utilizza il concetto di *equiprobabilità* degli eventi ancor prima di averla definita.

Nella pratica, invece, trova dei limiti laddove gli eventi non siano equiprobabili e, presupponendo un numero finito di risultati possibili, non trova applicazione nel continuo, ovvero nel caso in cui un esperimento possa restituire infiniti esiti [Erto et al., 2012].

L'approccio frequentista, che per molti aspetti ricalca lo schema della definizione classica, considera però la probabilità come risultato di esperimenti aleatori ripetuti infinite volte sotto le stesse condizioni.

In particolare, la probabilità che si verifichi un evento A , è calcolata come il limite cui tende la frequenza relativa di successo all'aumentare del numero di prove, in termini matematici:

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} \quad (\text{A.4})$$

La lettera n_A è associata al numero di volte che si verifica l'evento A , mentre la lettera n indica il numero totale delle prove; se l'esperimento è caratterizzato da una serie infinita di prove ripetute e indipendenti, l'equazione restituirà dunque la probabilità dell'evento oggetto di verifica [Agnoli, 2002].

Un esempio classico è quello del lancio della moneta non truccata. Si può constatare che, dopo un numero molto elevato di lanci, la frequenza relativa, ovvero la probabilità, di ottenere testa oppure croce, tenda al 50%.

Anche se in prima battuta potrebbe sembrare una definizione molto semplice e di banale applicazione, molte persone commettono errori senza neanche rendersene conto.

Il “paradosso di Monty Hall” chiamato anche “gioco delle tre scatole” rende bene l’idea: si immagina un gioco nel quale bisogna scegliere tra tre opzioni e solo una opzione è vincente. Dopo aver scelto un’opzione, il conduttore del gioco svela che una delle due rimanenti opzioni è perdente e propone al giocatore di cambiare l’opzione iniziale [Sanfilippo, 2015]. Cambiare l’opzione iniziale aumenterà la probabilità di vittoria del giocatore?

Molte persone risponderebbero che lo scambio non produrrebbe alcun vantaggio e che la probabilità di vittoria rimarrebbe il cinquanta per cento, non cambiando la scatola. In termini probabilistici, invece, converrebbe il cambio in quanto, così facendo, la probabilità di vittoria salirebbe a circa il 67%.

Basti pensare al fatto che la probabilità di aver scelto la scatola giusta all’inizio sarebbe pari a $\frac{1}{3}$, mentre la probabilità di aver scelto le scatole "sbagliate" sarebbe pari a $\frac{2}{3}$: il cambio dunque è sicuramente la miglior cosa da fare per cercare di massimizzare la probabilità di vittoria [Agnoli, 2002].

Un teorema che permetterebbe di dare una dimostrazione esaustiva è il *teorema di Bayes* visto in precedenza e pilastro di questo lavoro, poiché permette di aggiornare la probabilità ogni qualvolta entri in gioco una nuova informazione rilevante.

La probabilità frequentista è inoltre applicabile solo alle osservazioni e non permette di associare incertezza alle ipotesi; l’approccio bayesiano, invece, si libera dalle restrizioni legate alle frequenze, generalizzando il concetto di probabilità come “grado di fiducia” di un avvenimento qualsiasi e permette di applicare probabilità condizionate delle ipotesi a partire dalle probabilità condizionate delle varie osservazioni.

A.1.2.3 Approccio soggettivista

Come la definizione classica, anche quella frequentista non è esente da critiche. Nonostante superi alcuni limiti della definizione classica, l'approccio frequentista rimane fortemente limitante in quanto applicabile solo nel caso in cui l'esperimento sia replicabile infinite volte e sotto le stesse condizioni, due condizioni estremamente limitanti nella realtà.

Si pensi al caso in cui si vogliono fare previsioni sul successo di una missione spaziale, quando si vuole prevedere l'eruzione di un vulcano e tutti quei casi connessi alla vita dell'uomo nei quali le informazioni personali e gli stati d'animo sono di fondamentale importanza nei processi decisionali. Anche le decisioni finanziarie in caso di mercati estremamente volatili, oppure decisioni aziendali in condizioni di incertezza, ad esempio decisioni relative a effettuare investimenti appena finita l'emergenza sanitaria causata dal Coronavirus, sono casi in cui la concezione frequentista risulta poco utile .

Inoltre, le due concezioni della realtà precedentemente esposte, sono state qualificate come "oggettive" ma entrambe sono caratterizzate da componenti strettamente soggettive; basti pensare al concetto di equiprobabilità, di "stesse condizioni", di "infinito".

Per superare queste fragilità, all'inizio del XX secolo, è venuta a delinearsi una nuova corrente di pensiero, rappresentata dai "soggettivisti", di cui uno dei padri fondatori fu il matematico Bruno de Finetti.

Secondo tale approccio, la probabilità che si verifichi un certo evento A è intesa come la misura del *grado di fiducia* che un individuo *coerente* attribuisce al verificarsi dell'evento A , a seconda delle informazioni di cui dispone e delle opinioni [Agnoli, 2002].

Per misurare il *grado di fiducia* attribuito ad un evento, basti pensare ad una scommessa. Per poter godere del premio derivante da una scommessa, nel caso in cui l'evento sul quale abbiamo scommesso si verifichi, saremo disposti a pagare una determinata quota, chiamata anche *puntata*, per poter scommettere sull'evento dai noi ritenuto più probabile. Quanto più appaia verosimile tale evento, tanto più saremo disposti a pagare per poter partecipare alla scommessa; la somma che siamo disposti a pagare ben rappresenta il grado di fiducia contenuto nella definizione [De Finetti, 1931].

C'è un altro concetto fondamentale che merita di essere trattato, ovvero la nozione di *coerenza*.

In prima battuta, si potrebbe pensare che l'approccio soggettivista permetta piena libertà nell'assegnare una probabilità a qualsiasi evento ma, grazie alla limitazione della coerenza, non è così [Agnoli, 2002]. Tale regola che, contrariamente alla nozione di probabilità, è di natura oggettiva⁶, esclude tutte quelle valutazioni per le quali esista la probabilità di scommettere procurandosi un guadagno certo, comunque vadano le cose. Si vuole quindi evitare che le probabilità associate agli eventi vengano modificate qualora le informazioni in possesso non mutino, eliminando così la possibilità che uno scommettitore cambi la propria posizione per tornaconto personale.

Per esempio, uno scommettitore che, nel giudicare la probabilità di determinati eventi, contraddica un teorema del calcolo delle probabilità, non è coerente. Un "avversario" potrebbe scommettere con lui assicurandosi una vincita sicura oppure lo scommettitore in questione potrebbe utilizzare questa incoerenza volutamente per assicurarsi la vittoria [Crisma, 2006] [De Finetti, 1931].

Un esempio è sicuramente chiarificatore: si immagini che fra qualche giorno si disputerà la rivincita tra i due lottatori Chabib Abdulmanapovič Nurmagomedov e Conor Anthony McGregor per il titolo di campione del mondo pesi leggeri messo in palio dalla UFC, la più importante organizzazione a livello mondiale nel campo delle arti marziali miste. I fan sono scatenati e si precipitano nei centri scommesse di tutto il mondo per fare una puntata sul loro atleta preferito. Si immagini che l'allibratore⁷ in un centro scommesse offra la vittoria di Nurmagomedov a 3.2 e quella di McGregor a 1.25.

L'evento *Nurmagomedov vince* sarà indicato con A, mentre l'evento *McGregor vince* con B. Essendo B l'evento complementare di A, dovrà valere la relazione $P(A) + P(B) = 1$. La probabilità di A sarà pari a $\frac{1}{3.2}$ mentre la probabilità di B sarà $\frac{1}{1.25}$ poiché se si verifica l'evento A, $P(A) = 3.2 \cdot \frac{1}{3.2} = 1$, soddisfacendo così l'*assioma della certezza*; stesso ragionamento si farà per B.

Si può però notare che la somma delle probabilità dei due eventi è maggiore di uno, precisamente è pari a $\frac{1}{3.2} + \frac{1}{1.25} = 1.1125$. Questo sta a significare che l'allibratore si gioca in condizioni di vantaggio, visto che ha aumentato la probabilità complessiva che si verifichino i due eventi dell'11.25%. Si può infatti notare che l'allibratore, nel

⁶Ovvero consente di verificare se una valutazione di probabilità è coerente o meno usando gli strumenti della logica, senza dare importanza a come e a chi la fa.

⁷Chi allibra; in partic., chi registra le varie scommesse relative a competizioni sportive (spec. corse dei cavalli o dei cani), o anche elettorali, ecc., quando gli scommettitori sono numerosi; con questo sign., è usato anche in Italia il corrispondente termine ingl. bookmaker [Treccani, 1986].

caso in cui le puntate siano pari alle quote scambiate, incassi in totale 4,45 euro, dato dalla somma delle puntate, mentre spenderebbe 4 euro per ripagare gli scommettitori, valore ricavato dalla somma delle puntate rapportato alla probabilità che si realizzino, guadagnando 0,45 euro in ogni caso, sia che si realizzi A sia che si realizzi B.

Dall'esempio proposto, si evince un caso di violazione del concetto di coerenza, in quanto uno degli assiomi fondamentali del calcolo delle probabilità, l'assioma della certezza, non è stato rispettato, aprendo così all'opportunità di poter avere un guadagno certo.

La principale differenza tra questo approccio e un approccio oggettivista risiede, quindi, nel modo di concepire la probabilità: se l'approccio oggettivista la intende come una caratteristica intrinseca a un evento, l'approccio soggettivista la associa a una valutazione esterna sottostante a determinate regole.

Il vantaggio di questo modo di vedere la probabilità è il poter fare affermazioni probabilistiche su qualsiasi evento, anche quando essi non siano perfettamente equiprobabili, nel caso in cui l'esperimento non sia replicabile infinite volte oppure quando non sia possibile effettuarlo alle stesse condizioni [Agnoli, 2002].

B. Appendice

B.1 Variabili aleatorie

Dato lo *spazio di probabilità* (Ω, \mathcal{A}, P) e lo *spazio probabilizzante* (\mathbb{R}, β) , si definisce *variabile aleatoria* un'applicazione $X : \Omega \rightarrow \mathbb{R}$ tale che per ogni evento $\omega \in \Omega$ venga associato un numero reale x , il quale definisce la realizzazione della variabile casuale in ω . È quindi una funzione misurabile a valori reali definita su Ω che associa ad ogni esito di un esperimento casuale un certo numero reale; le variabili casuali sono strumenti per assegnare valori numerici agli *outcomes* di un esperimento. Con X si denota la variabile aleatoria in sé, mentre con x la sua realizzazione, il valore numerico specifico che potrebbe assumere.

Ci sono due macro categorie di variabili aleatorie: le variabili aleatorie discrete e le variabili aleatorie continue. Si parla di variabili aleatorie discrete quando i valori che assumono sono appunto discreti. Si pensi ad un esperimento volto a misurare l'altezza di un gruppo di persone in centimetri e arrotondare ciò che si è osservato a un numero intero. Nel caso di variabile aleatoria, riprendendo l'esempio appena fatto, basti immaginare a uno strumento di misura che è estremamente preciso e registra le altezze su un infinito numero di valori. In questo caso, come possibile realizzazione si potrebbe avere un qualunque numero reale; la nostra variabile aleatoria accetterebbe quindi valori sull'intero insieme dei numeri reali (in questo caso con l'esclusione dei numeri negativi), ovvero su un insieme continuo.

Il caso discreto è il caso più semplice e per questo motivo si presta bene a essere utilizzato per comprendere i concetti principali relativi alle variabili aleatorie, per poi adattarli al più articolato caso continuo [Durio and Isaia, 2009].

B.1.1 Variabili aleatorie discrete

In questa sezione tratteremo dunque le V.A. discrete ma anche i principali concetti per le variabili aleatorie in generale, in quanto il caso discreto risulta di più semplice applicazione e i risultati ottenuti possono essere facilmente adattati al caso continuo.

Come anticipato in precedenza, questa tipologia di V.A. assume valori nel discreto e ci interessa capire quanto queste realizzazioni siano probabili.

Si inizia quindi con il descrivere la variabile aleatoria attraverso un grafico a barre nel quale sulle ascisse vi sono le possibili realizzazioni, ossia le x , mentre sulle ordinate le $p_X(x)$. Per le varie realizzazioni, se ne calcolano le probabilità complessive sommando i risultati a esse associati nello spazio campionario e ciò genererà l'altezza della barra; maggiori sono i risultati nello spazio campionario associati a una determinata x , più la barra sarà alta.

L'altezza delle barre rappresenta la *probability mass function*, la PMF, in un particolare valore numerico x , la $p_X(x)$.

Matematicamente, la $p_X(x)$, è la probabilità che la nostra V.A. assuma il valore numerico x , e, per essere più precisi, $P(X = x)$, è la probabilità complessiva di tutti i risultati, indicati con ω , per i quali la variabile casuale finisce per assumere un determinato valore x , ovvero:

$$p_X(x) = P(X = x) = P(\omega \in \Omega : X(\omega) = x) \quad (\text{B.1})$$

Quindi, come detto in precedenza, l'altezza della barra è la probabilità complessiva di tutti gli ω collegati a una determinata realizzazione x .

Utile è ricordare che, essendo le PMF delle probabilità, esse non possono essere negative; inoltre, nel caso in cui si esauriscano tutti i valori delle realizzazioni, si esaurirebbero anche tutti i possibili risultati in Ω in quanto ogni risultato ω è collegato a una particolare realizzazione x . La somma delle diverse PMF deve dare come risultato 1, ossia $\sum_x p_X(x) = 1$ [Durio and Isaia, 2009].

Valore atteso di una variabile aleatoria. Il valore atteso è il concetto più importante quando si ha a che fare con le variabili aleatorie. Si può pensare come a una sorta di media, interpretando le probabilità come qualcosa di simile alle frequenze.

Un'associazione visiva può risultare utile per comprenderne il significato: pensando alla fisica, è un concetto molto vicino al concetto di *centro di gravità*, di *baricentro*, ovvero di quel punto che tiene in equilibrio un diagramma o una struttura.

Si pensi al seguente esempio: vi è un gioco nel quale, con $p = \frac{3}{6}$, vinci due euro, con $p = \frac{2}{6}$ vinci cinque euro e con $p = \frac{1}{6}$ ne vinci quattro.

Per calcolare il valore atteso basterà calcolare:

$$E[X] = \frac{3}{6} \times 2 + \frac{2}{6} \times 5 + \frac{1}{6} \times 4 = 3,33 \quad (\text{B.2})$$

Quindi, giocando infinite volte, ci si aspetta di vincere 3,33 euro.

Il valore atteso si è ottenuto, quindi, moltiplicando le probabilità associate ai diversi valori numerici per i rispettivi valori numerici, sommando poi i risultati ottenuti.

$$E[X] = \sum_x p_X(x)x \quad (\text{B.3})$$

Definito il valore atteso, un'analisi del suo comportamento e delle sue proprietà può risultare utile anche per gli argomenti successivi.

In *primis*, si pensi al caso in cui vi è una funzione Y della variabile aleatoria X . La variabile originaria X mappa i punti nella retta dei numeri reali, mentre Y è una funzione di questa V.A. e assume valori numerici completamente determinati una volta conosciuti i valori numerici di X . Sia X che Y sono, quindi, delle variabili aleatorie e, di conseguenza, è possibile calcolare anche il valore atteso di Y .

Per ogni particolare valore di y , ovvero per ogni realizzazione di Y , si collezionano tutti i possibili risultati associati a quel particolare valore di y e si trovano le rispettive probabilità, così pesate:

$$E[Y] = \sum_y p_Y(y)y \quad (\text{B.4})$$

C'è però un modo più semplice di fare lo stesso conteggio; sommando i valori sulla retta delle realizzazioni di X e non sommando i valori numerici sulle Y , otteniamo:

$$E[Y] = \sum_x g(x)p_X(x) \quad (\text{B.5})$$

Con \sum_x si intende la somma di tutti i risultati possibili, $p_X(x)$ ne restituisce la probabilità e $g(x)$ è quello che si osserva quando si verifica quella determinata x .

Anche se le due distinzioni possono apparire come una mera distinzione di stile, per calcolare le aspettative risultano molto differenti sul lato pratico; il secondo metodo ci permette infatti di calcolare il valore atteso di una funzione di una variabile aleatoria senza doverne ricavare la relativa PMF, semplificando così i calcoli e permettendo di lavorare con la PMF della funzione di origine, ossia della variabile aleatoria di partenza.

Può risultare anche utile specificare alcune proprietà delle aspettative:

- $E[\alpha] = \alpha$, con $\alpha \in \mathbb{R}$.
- $E[\alpha X] = \alpha E[X]$, grazie a questa proprietà è possibile estrarre una costante dal valore atteso senza creare distorsioni di alcun tipo.
- $E[\alpha X + \beta] = E[\alpha X] + \beta = \alpha E[X] + \beta$, con $\beta \in \mathbb{R}$: questa proprietà viene chiamata *linearità delle aspettative*, ovvero il valore atteso di una funzione lineare è pari alla funzione lineare del valore atteso.

Importante è rimarcare che le ultime due proprietà sono eccezioni della regola generale, secondo la quale il valore atteso della funzione di una variabile aleatoria è diverso dalla funzione del valore atteso della variabile aleatoria, $E[g(X)] \neq g(E[X])$ [Bertsekas and Tsitsiklis, 2008] [Durio and Isaia, 2009].

Varianza di una variabile aleatoria. Prima di introdurre questa nota caratteristica delle variabili aleatorie, è necessario inquadrare un concetto che ne permetterà la trattazione.

Quando si ha a che fare con una variabile aleatoria, potrebbe interessare conoscerne una sua funzione e, in questo caso, una funzione molto interessante è X^2 . Usando la regola generale del valore atteso avremo che:

$$E[X^2] = \sum_x x^2 p_X(x) \tag{B.6}$$

Nella pratica, questa funzione ci permette di calcolare la quantità che è di nostro interesse, ovvero la *varianza*, in quanto:

$$\text{var}(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 p_X(x) = E[X^2] - (E[X])^2 \quad (\text{B.7})$$

L'equazione B.6, restituisce il primo termine e, sottraendone il valore atteso della V.A. al quadrato, si ottiene la varianza, la quale viene anche indicata con σ_X^2 .

Concettualmente questa quantità serve per avere una visione più ampia di un determinato esperimento, spostando l'attenzione dal semplice valore atteso della distribuzione di probabilità che caratterizza quell'esperimento, al concetto di dispersione della distribuzione. La varianza è strutturata per dare maggiore enfasi alle distanze quadratiche tra i punti campionari e la media della V.A., enfatizzando così gli outliers. Questa quantità restituisce quanto, in media, le realizzazioni si allontanano dalla media. Anche la varianza gode di alcune proprietà:

- $\text{var}(X) \geq 0$, le varianze sono sempre quantità positive, in quanto sono somme di deviazioni quadrato.
- $\text{var}(\alpha X + \beta) = \alpha^2 \text{var}(X)$, si noti che, nel caso di una funzione lineare di una V.A., sommare una costante non influenza in alcun modo sulla dispersione, le distanze dalla media non subiscono nessun cambiamento. Tutto ciò che comporta sommare una costante β è uno spostamento in blocco a destra del diagramma e del relativo valore atteso, senza però influenzarne la variabilità; ecco perché si pone $\beta = 0$. Si noti anche che, quando si moltiplica una varianza per una generica costante α , la dispersione aumenta proporzionalmente al quadrato di quella costante.

La varianza presenta, però, un'unità di misura diversa di quella della variabile aleatoria di partenza, perché, come detto in precedenza, viene elevata al quadrato. Per ovviare a questo problema si ricorre allo scarto quadratico medio, comunemente chiamato *deviazione standard* o σ , calcolato applicando la radice quadrata alla varianza:

$$\sigma_X = \sqrt{\text{var}(X)} \quad (\text{B.8})$$

Questa quantità ci permette di trarre delle conclusioni sulla dispersione effettiva dei dati osservati, in quanto presenta la stessa unità di misura di questi ultimi e ne rende quindi possibile un confronto [Durio and Isaia, 2009].

Probabilità congiunte, marginali e condizionate di V.A. discrete multiple.

La trattazione delle *probabilità condizionate* necessita la conoscenza di alcuni concetti di base. Si immagini di avere a che fare con un determinato esperimento e che, in un dato momento, sopraggiungano nuove informazioni a riguardo. Questo evento comporterà un riadattamento del modello originale rappresentativo dell'esperimento, alla luce delle nuove informazioni disponibili.

Si passerà dunque da probabilità originali, relative al modello di partenza, a probabilità condizionate al verificarsi di un qualche nuovo evento. Quello che è interessante notare, è che le probabilità condizionate e quelle originali si comportano esattamente allo stesso modo, hanno le stesse proporzioni, semplicemente agiscono in un universo diverso. Le probabilità condizionate al verificarsi di un generico evento A , assumono la seguente forma:

$$p_{X|A}(x) = P(X = x|A) \tag{B.9}$$

L'unica operazione che bisogna fare quando si passa da probabilità originali a probabilità condizionate, è quella di ridimensionare le ultime in modo che la loro somma restituisca l'unità.

Si pensi a una variabile aleatoria che presenta una PMF Uniforme, ovvero tutte le possibili realizzazioni sono equiprobabili. Ipotizziamo che ci sia un gioco in cui i numeri 30, 40, 50 e 60 hanno tutti la stessa probabilità del 25% di verificarsi, qualcuno ci dice che il numero 30 non può verificarsi, l'evento condizionante sarà dunque $A = (x \geq 40)$. A questo punto le possibili realizzazioni sono soltanto tre e la loro distribuzione non varia, rimane sempre uniforme, i numeri 40, 50, 60 erano equiprobabili nell'universo iniziale e rimangono equiprobabili anche in quello condizionato. L'unica variazione da apportare è relativa ai valori delle PMF, in quanto non potranno rimanere invariate poiché la loro somma darebbe come risultato il 75% e sappiamo che la loro somma dovrebbe dare invece il 100%. Ciò che si farà, è semplicemente ridimensionarne le PMF in modo tale che la loro somma dia 1; a questo scopo, ogni realizzazione dovrà avere quindi una PMF pari a $\frac{1}{3}$.

Si può parlare anche di *valore atteso condizionato*, esso non presenta alcuna differenza dal valore atteso normale, se non quella che si applica sull'universo condizionato, servendosi delle PMF condizionate al posto delle PMF classiche:

$$E[X|A] = \sum_x p_{X|A}(x)x \quad (\text{B.10})$$

Analogamente, potremmo avere che:

$$E[g(X)|A] = \sum_x p_{X|A}(x)g(x) \quad (\text{B.11})$$

Si può notare che ogni formula incontrata in precedenza per le probabilità tradizionali, ha una controparte nell'universo condizionato.

Per la trattazione riguardante le *probabilità congiunte*, bisogna evidenziare il fatto che a un esperimento casuale si possono associare più variabili aleatorie. Per esempio, analizzando un gruppo di persone, potrebbe interessare sia l'età degli individui esaminati che il loro peso. Interessante è analizzare le due variabili congiuntamente, per catturare, attraverso le probabilità congiunte, una qualche associazione tra età e peso,

Queste probabilità hanno, dunque, lo scopo di catturare eventuali associazioni tra molteplici variabili casuali. Riferendosi al caso di due V.A., X e Y, la PMF congiunta, la probabilità che si verifichi una particolare coppia di risultati numerici è:

$$p_{X,Y}(x, y) = P(X = x, Y = y) \quad (\text{B.12})$$

Tipicamente, nel caso discreto, per rappresentare visivamente le PMF congiunte nel caso di due V.A., ci si avvale di una tabella posta sull'asse cartesiano, ove sulle ascisse ci sono i valori numerici che può assumere la variabile aleatoria X, e sulle ordinate i valori numerici per la variabile aleatoria Y; i valori scritti all'interno delle caselle che compongono questa tabella sono le probabilità congiunte.

Anche le probabilità congiunte, essendo delle probabilità, non possono assumere valori negativi e la somma di delle probabilità relative a tutte le possibili coppie di valori numerici che si possono ottenere deve essere pari a uno, ovvero $\sum_x \sum_y p_{X,Y}(x, y) = 1$. Conoscendo le probabilità congiunte, si potrebbe essere interessati solamente a una variabile, per esempio alla distribuzione della X. Avendo a disposizione le probabilità

congiunte si cerca di trovare la probabilità che X assuma un determinato valore x, utilizzando questa formula:

$$p_X(x) = \sum_y p_{X,Y}(x, y)^1 \quad (\text{B.13})$$

Questo tipo di probabilità si chiama *probabilità marginale* e si trova semplicemente fissando un particolare valore di X e sommando le PMF congiunte su tutti i possibili valori di Y, in quanto sono tutti i possibili modi con cui quel determinato valore di X si può verificare. Nel caso in cui si volesse trovare la probabilità di un valore di Y, vale lo stesso ragionamento: si fissa un determinato valore di Y e si sommano le PMF congiunte di tutte le X relative a quel valore specifico di Y.

A questo punto si può dire qualcosa in più rispetto alle *probabilità condizionate*. Esse sono definite come:

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p_{X|Y}(x, y)_2}{p_Y(y)} \quad (\text{B.14})$$

Matematicamente, questo significa normalizzare la PMF congiunta originale di un certo fattore $p_Y(y)$, cioè ridimensionarla di questo fattore, in modo che la somma delle probabilità condizionate sia uno.

In altri termini, la PMF condizionata, dato un particolare valore di Y, è una porzione della PMF congiunta, ha la stessa forma ma è ridimensionata in modo tale che la somma dei valori numerici che la compongono sia pari a uno.

Si noti che le probabilità marginali, condizionate e congiunte, sono correlate tra loro. In particolare:

$$p_{X,Y}(x, y) = p_X(x)p_{Y|X}(y|x)^3 \quad (\text{B.15})$$

Si noti infatti che la probabilità congiunta è pari alla probabilità marginale per la probabilità condizionata di Y su X.

¹Ovviamente vale anche il contrario, ovvero $p_Y(y) = \sum_x p_{X,Y}(x, y)$.

²Non è nulla di nuovo, è solamente un modo diverso di scrivere quanto detto già nella sezione sull'impostazione assiomatica, quando si parlava della probabilità di un evento A_1 condizionato da un evento A_2 .

³È semplicemente un modo diverso di scrivere la probabilità dell'intersezione di due eventi, $P(A \cap B) = P(A)P(B|A)$, pensando all'evento A come all'evento in cui X assume il valore x e all'evento B come a quell'evento in cui Y assume il valore y.

Queste nozioni sono molto utili poiché nel mondo reale un esperimento può essere caratterizzato da molteplici quantità incerte e bisogna essere in grado di analizzarle simultaneamente.

Ad esempio, nel caso di tre variabili casuali X, Y, Z avremo che la probabilità congiunta sarà:

$$p_{X,Y,Z}(x, y, z) = p_X(x)p_{y|X}(y|x)p_{z|X,Y}(z|x, y)^4 \quad (\text{B.16})$$

Mentre la pmf marginale di X sarà:

$$p_X(x) = \sum_{Y,Z} p_{X,Y,Z}(x, y, z) \quad (\text{B.17})$$

Anche in questo caso fissiamo una determinata $X = x$ e consideriamo tutte le probabilità per le restanti variabili Y e Z , sommando come fatto in precedenza queste probabilità, otteniamo la PMF marginale di X [Bertsekas and Tsitsiklis, 2008].

Concetto di indipendenza con variabili aleatorie discrete multiple. Due eventi si dicono indipendenti se l'avverarsi del primo non incide sul verificarsi del secondo, la probabilità condizionata di A dato B è data da:

$$P(A|B) = P(A) \quad (\text{B.18})$$

Pertanto, la probabilità che si verifichino sia l'evento A che l'evento B sarà:

$$P(A \cap B)P(A|B)P(B) = P(A)P(B) \quad (\text{B.19})$$

Due eventi sono indipendenti se la probabilità che si verifichino entrambi è pari al prodotto delle relative probabilità.

Il caso di tre eventi è un po' più articolato ma, fortunatamente, nel generalizzare questo concetto per le variabili aleatorie, le cose si fanno più semplici.

Tre o più variabili aleatorie si dicono indipendenti se il verificarsi di una non dipende

⁴ $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$, si noti il parallelismo fra le due formule.

dal verificarsi delle altre. Nel caso di X, Y, Z avremo quindi che:

$$p_{X|Y,Z}(x|y, z) = p_X(x)^5 \quad (\text{B.20})$$

Come in precedenza, la PMF congiunta di più variabili casuali è pari al prodotto delle relative probabilità. Per esempio, nel caso di X, Y e Z avremo che, $\forall x, y, z$:

$$p_{X,Y,Z}(x, y, z) = p_X(x)p_Y(y)p_Z(z) \quad (\text{B.21})$$

Dietro queste formule c'è quanto detto per il caso degli eventi, ovvero che conoscere alcune realizzazioni di una determinata V.A. non influenza le convinzioni originali sulle altre realizzazioni. Le credenze a priori non variano in alcun modo con l'aggiunta di informazioni e quindi combaciano con le credenze a posteriori. Ciò comporta che il prodotto delle probabilità relative alle singole V.A. sia pari alla probabilità che tutte le V.A. si verifichino congiuntamente [Bertsekas and Tsitsiklis, 2008].

Valore atteso di variabili aleatorie multiple discrete. Avendo introdotto la possibilità di lavorare con più di una variabile aleatoria, è interessante notare come si comporta il valore atteso in questo caso. Generalizzando l'equazione B.3 avremo che:

$$E[g(X, Y)] = \sum_x \sum_y p_{X,Y}(x, y)g(x, y) \quad (\text{B.22})$$

Quando si ha a che fare con funzioni di più variabili aleatorie, il valore atteso di queste funzioni sarà diverso dalla funzione dei rispettivi valori attesi, ovvero:

$$E[g(X, Y)] \neq g(E[X], E[Y]) \quad (\text{B.23})$$

Come già visto per il caso del valore atteso di una singola variabile aleatoria, questa regola generale trova l'eccezione quando vi è una funzione lineare delle aspettative, ma quando si ha a che fare con variabili casuali multiple vale anche la seguente proprietà:

$$E[X + Y + Z] = E[X] + E[Y] + E[Z] \quad (\text{B.24})$$

⁵Solo se $p_Y(y) > 0$ e $p_Z(z) > 0$, in quanto le probabilità condizionate sono definite solamente per probabilità positive.

Risulta quindi che il valore atteso si comporta in maniera lineare, ovvero il valore atteso della somma di più variabili aleatorie è pari alla somma dei valori attesi delle diverse variabili aleatorie.

Un' altra grande eccezione alla regola generale si ha quando le variabili aleatorie sono tra loro indipendenti, in questo caso si avrebbe che:

$$E[XY] = E[X]E[Y]^6 \quad (\text{B.25})$$

Si dimostra in [Bertsekas and Tsitsiklis, 2008] che il valore atteso del prodotto di V.A. indipendenti è pari al prodotto dei valori attesi.

Varianza di variabili aleatorie multiple discrete. Oltre alle proprietà della varianza già viste per il caso di una singola variabile casuale, quando si sommano due o più V.A. potrebbe interessare capire il comportamento delle relative varianze, cioè se si comportano in modo lineare come il valore atteso.

Fortunatamente, sotto la condizione di indipendenza, anche le varianze presentano le buone proprietà rese possibili dalla linearità. Infatti, se abbiamo che:

$$Z = X + Y, \quad (\text{B.26})$$

e se X e Y sono indipendenti, allora:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) \quad (\text{B.27})$$

Per far capire l'importanza del requisito di indipendenza, è utile guardare ai due casi limite di estrema dipendenza:

1. $X = Y \rightarrow \text{var}(X + Y) = \text{var}(2Y) = 4 \cdot \text{var}(x)$
2. $X = -Y \rightarrow \text{var}(X + Y) = \text{var}(0) = 0$

In entrambi i casi si può notare come i risultati non combaciano con i risultati che restituirebbe la formula B.27 nel caso di indipendenza [Bertsekas and Tsitsiklis, 2008].

⁶Questa relazione è valida anche nel caso del valore atteso del prodotto di funzioni delle V.A. originali.

B.1.2 Variabili aleatorie continue

Le variabili aleatorie continue prendono valori nel continuo, i valori assunti dalla V.A. possono essere qualsiasi numero reale. Inoltre, a differenza delle variabili aleatorie discrete, quelle continue sono descritte dalla *probability density function*, abbreviata PDF, che indicheremo con f_X . Prendendo valori su tutto l'asse dei numeri reali, la PDF si disperde lungo tutto l'asse reale generando in questo modo una funzione continua.

Importante è sapere che, nel caso continuo, l'area sottostante alla curva generata dalla V.A corrisponde al valore della probabilità e si calcola attraverso l'uso degli integrali.

Proprio per questo motivo, le probabilità che la variabile assuma un valore specifico saranno nulle: integrare l'area sotto una curva relativamente a un punto, restituisce un *output* pari a zero [Durio and Isaia, 2009] [Bertsekas and Tsitsiklis, 2008].

Quando si calcola la probabilità nel caso continuo bisogna ragionare sempre in termini di intervalli. Infatti, una variabile aleatoria si dice continua se si possono calcolare le probabilità in questo modo:

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx \quad (\text{B.28})$$

Si noti che nel caso discreto questa formula sarebbe:

$$P(a \leq X \leq b) = \sum_{x \in [a,b]} p_X(x) \quad (\text{B.29})$$

Da questo parallelismo, si può notare che le somme nel caso discreto si trasformano in integrali nel caso continuo e che le PMF, le $p_X(x)$, si trasformano in PDF, $f_X(x)$. Questo aspetto è di estrema rilevanza perché si può generalizzare per tutte le formule viste nel caso discreto.

Avendo introdotto il concetto di PDF, utile è comprenderne le principali proprietà:

1. $f_X \geq 0$: le densità, essendo utilizzate per il calcolo delle probabilità, non possono assumere valori negativi;
2. $\int_{-\infty}^{\infty} f_X(x) d(x) = 1$: la probabilità totale generata dall'insieme di tutti i valori reali, quindi l'area sottostante alla funzione, deve essere pari a uno.

Ora è utile comprenderne il comportamento e, per questo fine, risulta utile ragionare in termini di intervalli piccolissimi, di lunghezza δ . Avremo:

$$P(x \leq X \leq x + \delta) = \int_x^{x+\delta} f_X(x)dx \approx f_X(x)dx\delta \quad (\text{B.30})$$

Quindi, la probabilità di un piccolo intervallo sarà pari alla densità per δ , ovvero l'altezza per la base dell'intervallo.

Isolando $f_X(x)$ nella formula B.30, otteniamo:

$$f_X(x) = \frac{P(x \leq X \leq x + \delta)}{\delta} \quad (\text{B.31})$$

Si è fatto questo passaggio per far comprendere che le densità non sono probabilità pure, ma sono probabilità per unità di lunghezza e, per questo motivo, possono essere anche maggiori di uno.⁷

A questo punto, con la precisazione fatta in precedenza sulle formule nel caso continuo, se ne elencano le principali:

1. $E[X] = \int_{-\infty}^{\infty} f_X(x)d(x)x$, è la formula per il valore atteso nel caso continuo;
2. $E[g(X)] = \int_{-\infty}^{\infty} f_X(x)d(x)g(x)$, ovvero il valore atteso di una funzione di una V.A. continua;
3. $\sigma_X^2 = var(x) = \int_{-\infty}^{\infty} f_X(x)d(x)(X - E(X))^2$.

Cumulative distribution function. Si è visto che ci sono formule per il caso discreto e formule per il caso continuo, interessante sarebbe trovare un modo per gestire le due tipologie di variabili con uno strumento comune, che si applichi sia a V.A. discrete che continue.

Fortunatamente questo strumento esiste e si chiama *cumulative distribution function*, abbreviato CDF.

La CDF è la probabilità che la V.A. X assuma valori minori di una determinata x :

$$F_X(x) = P(X \leq x) \quad (\text{B.32})$$

⁷Sempre con il vincolo che l'area totale sotto la curva sia 1.

Per il caso discreto avremo:

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k), \quad (\text{B.33})$$

mentre per il caso continuo:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt. \quad (\text{B.34})$$

In tutti e due i casi le CDF saranno ben definite. Nel caso continuo sarà una funzione continua che cresce in maniera continua da zero fino a stabilizzarsi a uno, in quanto alimentata dall'accumularsi delle probabilità muovendosi da sinistra verso destra.

Nel caso discreto la crescita della CDF è leggermente diversa; anch'essa è crescente da zero a uno, ma non cresce in maniera lineare, segue una funzione a tratti con i salti in corrispondenza dei punti campionari, salti determinati nel loro ammontare dalle masse di probabilità di quei determinati punti.

Questa tecnica è molto funzionale sul lato pratico perché permette, grazie alla sua flessibilità, di calcolare le probabilità di V.A. che magari non sono propriamente né continue né discrete, ma sono V.A. miste. Trattare questo tipo di variabili con le PMF o le PDF potrebbe essere molto complicato, mentre l'utilizzo delle CDF semplifica di molto la procedura [Durio and Isaia, 2009] [Bertsekas and Tsitsiklis, 2008].

Probabilità congiunte, marginali e condizionate di V.A. continue multiple.

Come nel caso discreto, anche nel caso continuo nella maggior parte degli esperimenti ci si dovrà confrontare con più variabili aleatorie e, conseguentemente, con i concetti relativi alle probabilità congiunte, marginali e condizionate [Bertsekas and Tsitsiklis, 2008] [Tsitsiklis, 2010]. Iniziamo con la formula per due variabili che si verificano congiuntamente:

$$P((X, Y) \in S) = \int \int_S f_{X,Y}(x, y) dx dy \quad (\text{B.35})$$

Due variabili si dicono continue e congiunte se è possibile calcolare la probabilità integrando una determinata funzione chiamata *funzione di densità congiunta* sull'insieme d'interesse.

In questo caso, bisogna ragionare pensando ad uno spazio bidimensionale. In effetti, la densità congiunta bisogna immaginarla, nel caso di due variabili, come una sorta di superficie relativa all'evento S di cui ci interessa calcolare la probabilità.

Una volta a disposizione la densità congiunta, si procede pensando a questa come a una sorta di superficie all'interno del piano bidimensionale; la probabilità di cadere nell'insieme S la si trova calcolando il volume sottostante questa superficie. Essendo i volumi associati alle probabilità, il volume totale sottostante all'area S sarà pari a uno:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y} dx dy = 1 \quad (\text{B.36})$$

e, poiché stiamo parlando di probabilità, la densità dovrebbe essere una funzione positiva:

$$f_{X,Y} \geq 0 \quad (\text{B.37})$$

Come per il caso di una variabile aleatoria continua, anche qui ragioniamo in termini di piccoli intervalli; ovviamente, avendo due variabili, si ragionerà su due piccoli intervalli. Immaginando il piano bidimensionale, si immagini di prendere un piccolo intervallo di lunghezza δ sulle x e un piccolo intervallo δ sulle y , ovvero si guardi all'evento in cui X cade nell'intervallo $[x, x + \delta]$ e simultaneamente Y cade nell'intervallo $[y, y + \delta]$. L'evento potrebbe risultare un rettangolo o un quadrato molto piccolo, il cui volume sottostante a tale quadrato sarà quindi pari alla sua area base, δ^2 , per l'altezza, ovvero il valore della funzione in corrispondenza di quella determinata area. Quindi:

$$P(x \leq X \leq x + \delta, y \leq Y \leq y + \delta) \approx f_{X,Y}(x, y) \delta^2 \quad (\text{B.38})$$

La formula mette in evidenza come le probabilità congiunte siano delle densità per un'area e, se isolassimo la densità congiunta, si noti che quest'ultima restituisce un probabilità relativa a una determinata area.

Si supponga di avere una funzione di due variabili aleatorie descritte da una densità congiunta, quello che si può fare è trovarne il valore atteso, ovvero:

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy \quad (\text{B.39})$$

Similmente a quanto fatto per le variabili discrete multiple, anche in questo caso può interessare derivare la *probabilità marginale* di una delle variabili partendo dalla *probabilità congiunta* di tutte le variabili in gioco.

Si immagini ancora il caso più semplice, ovvero il caso con due variabili aleatorie X e Y. Si pensi ad un piano cartesiano che sulle ascisse presenta i valori di X e sulle ordinate in valori di Y; cercando di ricavare la probabilità marginale relativa a un piccolo intervallo di valori delle x , si avrà:

$$P(x \geq X \geq x + \delta) = f_X(x)\delta \quad (\text{B.40})$$

dove $f_X(x)\delta$ è la densità marginale; quello che però ci interessa è esprimerla in termini di probabilità congiunta. La formula è la seguente:

$$P(x \geq X \geq x + \delta) = f_X(x)\delta = \int_{-\infty}^{\infty} \int_x^{x+\delta} f_{X,Y}(\bar{x}, y) d\bar{x} dy \quad (\text{B.41})$$

Si sta integrando da meno infinito a più infinito per quanto riguarda Y e da x a $x+\delta$ per quanto riguarda X, quindi stiamo integrando ogni y sulla dimensione rappresentata dal piccolo intervallo d'interesse δ per risolvere l'integrale di X e poi si integra sulla dimensione di Y per l'integrale rimanente.

L'integrale relativo alla X varia in maniera infinitesimale, motivo per cui può essere considerato come una costante. In questo modo le cose si semplificano e si avrà:

$$f_X(x)\delta = \int_{-\infty}^{\infty} \delta f(x, y) dy \quad (\text{B.42})$$

Semplificando i δ :

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (\text{B.43})$$

Questa formula non dovrebbe sorprendere, è esattamente come per il caso discreto presentando solamente integrali al posto delle somme e la PDF congiunta al posto della PMF congiunta.

Infine, vi sono le *probabilità condizionate*; nel caso continuo i concetti sono un po' più sottili rispetto al caso discreto. Le densità condizionate sono come le densità classiche, ovvero probabilità relative a piccoli intervalli, ma all'interno di un nuovo universo condizionato. Per universo condizionato si intende un universo nel quale sono note

informazioni riguardanti le altre variabili aleatorie.

In termini matematici:

$$P(x \leq X \leq x + \delta | Y \approx y) \approx f_{X|Y}(x|y)\delta \quad (\text{B.44})$$

Ecco la sottigliezza: si condiziona su un evento in cui Y è molto vicino a y poiché, se si fosse utilizzata l'uguaglianza, nel caso continuo la probabilità di un punto è nulla e condizionare su un evento nullo non è consentito. In questo modo invece si condiziona su un evento in cui Y è estremamente vicina a y ma presentando una probabilità positiva, così da avere le caratteristiche per poter essere un evento condizionante.

La quantità di maggiore interesse è la densità condizionata, pari a:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad (\text{B.45})$$

A livello interpretativo, è utile pensare di fissare la y , cioè il valore della V.A. Y , e si osserva la densità condizionata in funzione di X , la quale avrà la stessa forma della PDF congiunta.

La PDF condizionata non è nient'altro che una sezione della PDF congiunta e il termine $f_Y(y)$ serve proprio a ridimensionare la congiunta e a cambiarne forma.

Nel caso di indipendenza⁸, la densità congiunta sarà pari al prodotto delle marginali, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ e, sostituendo questo prodotto all'interno del numeratore dell'equazione B.45, si può notare che la densità condizionata sarà pari alla marginale dell'evento condizionato. Questo risultato è in linea con l'idea generale di condizionamento sotto l'ipotesi d'indipendenza, ovvero che avere informazioni sull'evento condizionante non altera le credenze sull'evento condizionato.

Nel caso di indipendenza tra due V.A., si avrà che:

$$f_{X|Y}(x|y) = f_X(x) \quad (\text{B.46})$$

Una prospettiva 3D potrebbe aiutare a comprendere ancora meglio il tema delle distribuzioni congiunte, marginali e condizionate.

⁸Non varia concettualmente dal caso discreto, varia solo la notazione.

Joint Density

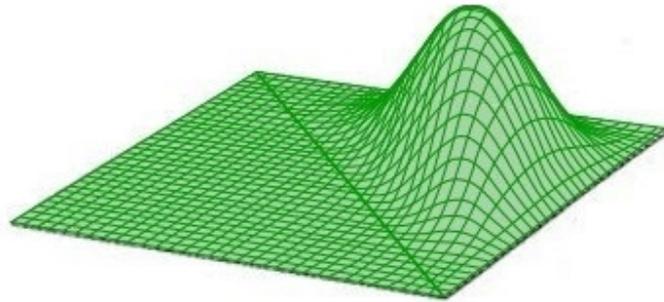


Figura B.1: Image by MIT OpenCourseWare, adapted from Probability, by J. Pittman, 1999.

Questa potrebbe essere un esempio di PDF congiunta. Per passare alla PDF marginale, come detto in precedenza, si seziona l'immagine fissando una delle due variabili e il risultato è descritto dalla figura B.2. L'area totale sotto ogni singola sezione, restituisce

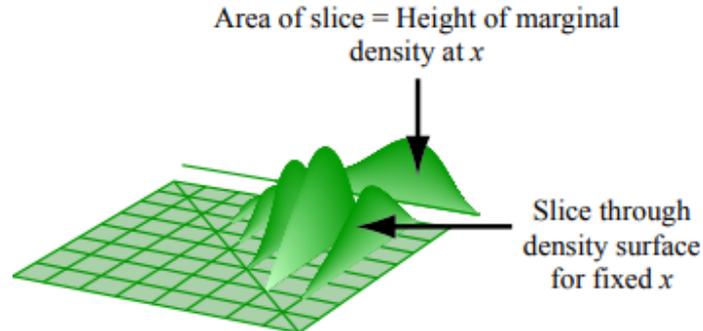


Figura B.2: Image by MIT OpenCourseWare, adapted from Probability, by J. Pittman, 1999.

la densità marginale delle rispettive x ; si fissa una particolare x e si integra su tutte le y relative, ciò restituirà densità marginale per quella determinata x .

Per ricavare la densità condizionata, per esempio $f_{Y|X}(y|x)$, bisogna ragionare in questo modo: sapendo che il condizionamento sarà sulle x , significa che ci si sposta da un universo nel quale non abbiamo informazioni ad un universo in cui si ha la conoscenza

che la X ha assunto un particolare valore x . Ci si sposta quindi nell'universo relativo all'evento $X = x$ e ci si focalizza, come si può notare dall'immagine sopra, su quella sezione relativa a quella particolare x . A questo punto, la probabilità totale delle diverse y relative a quella determinata x , dovrebbe essere pari a 1. Per fare questo, si normalizza la densità congiunta dividendola per la densità marginale, in questo caso per la densità marginale di x .

Se facessimo questa operazione condizionando su tutti le realizzazioni di x , otterremmo qualcosa di simile a questo grafico:

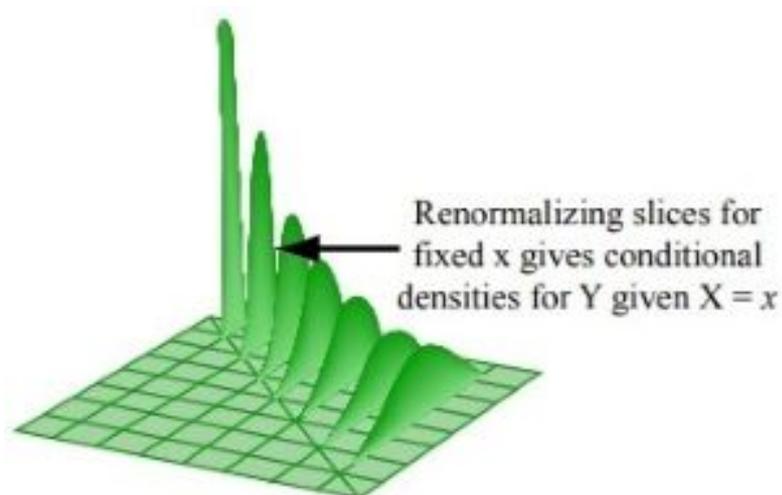


Figura B.3: Image by MIT OpenCourseWare, adapted from Probability, by J. Pittman, 1999.

C. Appendice

C.1 Codice R

Si riporta di seguito il codice R scritto in collaborazione con la Professoressa Nava:

```
# import
#input Stata file
install.packages("foreign")
library(foreign)
mydata <- read.dta("ESS9e02.dta")
# visualizzo nome delle variabili
names(mydata)
# centry nome dello Stato
table(mydata$centry)
# lista partiti votati in italia
table(mydata$prvtvcit)
# ricodifico in base ai Gruppi di partito europei, associo al partito "X" la relativa
famiglia europea "Y"
votoit<-mydata$prvtvcit
levels(votoit)
levels(votoit)<-c("S&D","ADLE","NON DISPONIBILE","PPE","PPE","GUE/NGL",
"EFDD","PPE","ENL","ECR","PPE","GUE/NGL","NON DISPONIBILE","Altro",
"Not applicable","Refusal","Don't know","No answer")
levels(votoit)
```

```

table(votoit)
# Francia
table(mydata$prtvtdfr)
# ricodifico in base a Gruppi
votofr<-mydata$prtvtdfr
levels(votofr)
levels(votofr)<-c("GUE/NGL","GUE/NGL","GUE/NGL","GUE/NGL","GUE/NGL",
"VERDI/ALE","ADLE","ADLE","PPE","EFDD","ENL", "Altro","Not applicable",
"Refusal","Don't know","No answer")
levels(votofr)
# Bulgaria
votobg<-mydata$prvtbdbg
levels(votobg)
levels(votobg)<-c("PPE","S&D","ECR","ADLE","NON DISPONIBILE","PPE","S&D",
"NON DISPONIBILE","PPE","NON DISPONIBILE","NON DISPONIBILE", "Al-
tro","Invalid", "Not applicable","Refusal","Don't know","No answer")
levels(votobg)
# Austria
votoat<-mydata$prvtvcat
levels
levels(votoat)<-c("S&D","PPE","ENL","VERDI/ALE","VERDI/ALE","GUE/NGL",
"ADLE", "NON DISPONIBILE", "Altro","Not applicable","Refusal","Don't know","No
answer")
levels(votoat)
# Belgio
votobe<-mydata$prvtvdbbe
levels(votobe)
levels(votobe)<-c("VERDI/ALE","PPE","ECR","ECR","S&D","GUE/NGL","ENL",
"ADLE","PPE","VERDI/ALE","NON DISPONIBILE","ADLE","S&D","GUE/NGL",

```

```

"NON DISPONIBILE", "NON DISPONIBILE", "Altro","BlanK", "Invalid", "Not
applicable","Refusal","Don't know","No answer")
levels(votobe)
# Svizzera (anche se non appartiene all'UE, è stato possibile associare ai vari partiti
norvegesi famiglie di partito europee)
votoch<-mydata$prvtvtgch
levels(votoch)
levels(votoch)<-c("NON DISPONIBILE","S&D","ADLE","PPE","VERDI/ALE","ADLE",
"NON DISPONIBILE", "ECR","ECR","NON DISPONIBILE", "GUE/NGL", "NON
DISPONIBILE","GUE/NGL","VERDI/ALE", "Altro","BlanK", "Not applicable","Refusal",
"Don't know","No answer")
levels(votoch)
# Cipro
votocy<-mydata$prvtvtbcy
levels(votocy)
levels(votocy)<- c("GUE/NGL","S&D","PPE","NON DISPONIBILE","ECR","VERDI/ALE",
"S&D", "S&D", "Altro", "Not applicable","Refusal","Don't know","No answer")
levels(votocy)
# Germania 1
votode1<-mydata$prtvede1
levels(votode1)
levels(votode1)<-c("PPE","S&D","GUE/NGL","VERDI/ALE","ADLE","EFDD",
"VERDI/ALE", "NON DISPONIBILE", "Altro", "Not applicable","Refusal","Don't
know","No answer")
levels(votode1)
# Germania 2
votode2<-mydata$prtvede2
levels(votode2)
levels(votode2)<-c("PPE","S&D","GUE/NGL","VERDI/ALE","ADLE","EFDD",

```

```

"VERDI/ALE", "NON DISPONIBILE", "Altro", "Not applicable","Refusal","Don't
know","No answer") levels(votode2)
# Estonia
votoee<-mydata$prvtvtee
levels(votoee)
levels(votoee)<-c("ADLE","ADLE","PPE","S&D","VERDI/ALE","ENL","NON DI-
SPONIBILE", "NON DISPONIBILE", "ECR", "NON DISPONIBILE", "Altro", "Not
applicable","Refusal","Don't know","No answer")
levels(votoee)
# Spagna
votoes<-mydata$prvtvtees
levels(votoes)
levels(votoes)<-c("PPE","S&D","GUE/NGL","GUE/NGL","ADLE","GUE/NGL","NON
DISPONIBILE", "NON DISPONIBILE", "VERDI/ALE", "VERDI/ALE","ADLE",
"VERDI/ALE", "ADLE","GUE/NGL", "S&D","ECR","NON DISPONIBILE", "VER-
DI/ALE","NON DISPONILE","GUE/NGL", "Altro","Blank", "Invalid", "Not appli-
cable","Refusal","Don't know","No answer")
levels(votoes)
# Finlandia
votofi<-mydata$prvtvtdfi
levels(votofi)
levels(votofi)<-c("PPE","ADLE","ADLE","ECR","PPE","NON DISPONIBILE",
"VERDI/ALE", "NON DISPONIBILE", "NON DISPONIBILE", "VERDI/ALE","S&D",
"GUE/NGL", "NON DISPONIBILE","NON DISPONIBILE", "NON DISPONIBILE",
"Altro", "Not applicable","Refusal","Don't know","No answer")
levels(votofi)
# UK
votogb<-mydata$prvtvcgb
levels(votogb)
levels(votogb)<-c("PPE","S&D","ADLE","VERDI/ALE","VERDI/ALE","VERDI/ALE",

```

```

"ENL", "Altro", "ECR", "NON DISPONIBILE", "GUE/NGL", "EFDD", "ADLE",
"S&D", "VERDI/ALE", "NON DISPONIBILE", "GUE/NGL", "NON DISPONIBI-
LE", "Altro", "Not applicable", "Refusal", "Don't know", "No answer")
levels(votogb)
# Croazia
votohr<-mydata$prvtvtahr
levels(votohr)
levels(votohr)<-c("PPE", "S&D", "ECR", "NON DISPONIBILE", "NON DISPONIBILE",
"NON DISPONIBILE", "NON DISPONIBILE", "S&D", "ADLE", "Altro", "Not ap-
plicable", "Refusal", "Don't know", "No answer")
levels(votohr)
# Ungheria
votohu<-mydata$prvtvtfhu
levels(votohu)
levels(votohu)<-c("S&D", "VERDI/ALE", "PPE", "NON DISPONIBILE", "PPE",
"VERDI/ALE", "S&D", "NON DISPONIBILE", "ADLE", "VERDI/ALE", "NON DI-
SPONIBILE", "Altro", "Not applicable", "Refusal", "Don't know", "No answer")
levels(votohu)
# Iralanda
votoie<-mydata$prvtvcie
levels(votoie)
levels(votoie)<-c("GUE/NGL", "ADLE", "PPE", "VERDI/ALE", "NON DISPONIBILE",
"S&D", "GUE/NGL", "S&D", "GUE/NGL", "Altro", "Not applicable", "Refusal", "Don't
know", "No answer")
levels(votoie)
# Olanda
votonl<-mydata$prvtvgnl
levels(votonl)
levels(votonl)<-c("ADLE", "S&D", "ENL", "GUE/NGL", "PPE", "ADLE", "ECR", "VER-
DI/ALE", "ECR", "GUE/NGL", "NON DISPONIBILE", "NON DISPONIBILE", "ECR",

```

```

"NON DISPONIBILE", "Altro", "Blank", "Not applicable","Refusal", "Don't know", "No
answer")
levels(votonl)
# Norvegia (stesso discorso della Svizzera)
votono<-mydata$prvtvtno
levels(votono)
levels(votono)<-c("GUE/NGL","VERDI/ALE","S&D","ADLE","PPE","ADLE", "PPE",
"NON DISPONIBILE","NON DISPONIBILE", "VERDI/ALE", "Altro", "Not appli-
cable","Refusal", "Don't know", "No answer")
levels(votono)
# Polonia
votopl<-mydata$prvtvtdpl
levels(votopl)
levels(votopl)<-c("ECR","NON DISPONIBILE","NON DISPONIBILE","PPE","PPE",
"ECR", "S&D", "NON DISPONIBILE", "Altro", "Not applicable","Refusal","Don't
know","No answer")
levels(votopl)
# Portogallo
votopt<-mydata$prvtvcpt
levels(votopt)
levels(votopt)<-c("NON DISPONIBILE","NGL","NGL","NON DISPONIBILE","NON
DISPONIBILE","S&D", "NON DISPONIBILE", "NON DISPONIBILE", "PPE", "AD-
LE", "NON DISPONIBILE", "NON DISPONIBILE", "S&D","NON DISPONIBILE",
"VERDI/ALE","PPE", "Blank", "Altro", "Not applicable","Refusal", "Don't know", "No
answer")
levels(votopt)
# Serbia (stesso discorso della Svizzera e della Norvegia)
votors<-mydata$prvtvtrs
levels(votors)

```

```

levels(votors)<-c("NON DISPONIBILE","ADLE","NON DISPONIBILE","ECR","NON
DISPONIBILE", "NON DISPONIBILE", "NON DISPONIBILE", "NON DISPONIBI-
LE", "NON DISPONIBILE", "NON DISPONIBILE","PPE", "S&D", "Altro", "Not
applicable","Refusal","Don't know","No answer")
levels(votors)
# Svezia
votose<-mydata$prvtvtcse
levels(votose)
levels(votose)<-c("ADLE","ADLE","PPE","VERDI/ALE","PPE", "S&D", "GUE/NGL",
"S&D", "ECR", "NON DISPONIBILE", "Not applicable", "Refusal", "Don't know", "No
answer")
levels(votose)
# Slovenia votosi<-mydata$prvtvtfsi
levels(votosi)
levels(votosi)<-c("ADLE","GUE/NGL","ADLE","PPE","NON DISPONIBILE", "S&D",
"ADLE", "PPE", "PPE","ADLE", "NON DISPONIBILE", "Altro", "Not applica-
ble","Refusal", "Don't know", "No answer")
levels(votosi)
# Slovacchia
votosk<-mydata$prvtvtdsk
levels(votosk)
levels(votosk)<-c("ECR", "NON DISPONIBILE", "S&D", "NON DISPONIBILE", "NON
DISPONIBILE", "ECR", "NON DISPONIBILE", "NON DISPONIBILE", "Altro",
"Not applicable", "Refusal", "Don't know", "No answer") levels(votosk)
# creo un unico vettore per il voto
# install.packages("dplyr") and ("tidyverse")
library("dplyr")
install.packages("dplyr")
install.packages("tidyverse")
# vettore unico per il voto

```

```

voto<-coalesce(as.character(votoit), as.character(votofr), as.character(votobg),
as.character(votoat), as.character(votobe), as.character(votoch), as.character(votocy),
as.character(votode1), as.character(votode2), as.character(votoee), as.character(votoes),
as.character(votofi), as.character(votogb), as.character(votohr), as.character(votohu),
as.character(votoie), as.character(votonl), as.character(votono), as.character(votopl),
as.character(votopt), as.character(votors), as.character(votose), as.character(votosi),
as.character(votosk))
# MCMCGLMM
install.packages("MCMCglmm")
library(MCMCglmm)
# id missing da rimuovere
id_rm<-which(!complete.cases(voto))
# uso solo persone con voto palese e non NA
mydata1<-mydata[-id_rm,]
mydata1$voto<-voto[-id_rm]
# creo variabile con id del choiceset
# importo file XLSX
library(openxlsx)
gruppi<-read.xlsx("gruppi.xlsx")
# tolgo missing che non mi servono e seconda variabile e rinONimo la terza
gruppi<-gruppi[which(complete.cases(gruppi$Country)),c(1,3)]
names(gruppi)<-c("Country", "Group")
# assegno in mydata1 una variabile contenente l'id del gruppo
mydata1$group<-rep(0,dim(mydata1)[1]) for (i in unique(mydata1$cntry)) mydata1$group
[which (mydata1$cntry==i)]<-gruppi$Group [which(gruppi$Country==i)]
# da mydata1 tolgo anche i livelli di voto residuali che non mi interessano
unique(mydata1$voto)
# Invalid
# Blank

```

```

# NON DISPONIBILE
# NON DISPONILE
# BlanK
id_rm<-which(mydata1$voto=="Invalid")
id_rm<-union(id_rm, which(mydata1$voto=="Blank"))
id_rm<-union(id_rm, which(mydata1$voto=="NON DISPONIBILE"))
id_rm<-union(id_rm, which(mydata1$voto=="NON DISPONILE"))
id_rm<-union(id_rm, which(mydata1$voto=="BlanK"))
length(id_rm) # perdo 2114 osservazioni
mydata2<-mydata1[-id_rm,]
dim(mydata1)
dim(mydata2)
# tolgo missing dalle variabili che voglio utilizzare e ridefinisco i livelli delle variabili
mydata2<-mydata2[-which(!complete.cases(mydata2$agea)),]
dim(mydata2)
mydata2<-mydata2[-which(!complete.cases(mydata2$eisced)),]
dim(mydata2)
mydata2<-mydata2[-which(!complete.cases(mydata2$mbtru)),]
dim(mydata2)
levels(mydata2$mbtru)<-c("Yes, currently", "Yes, previously", "No", "No", "No", "No")
mydata2<-mydata2[-which(!complete.cases(mydata2$hinctnta)),]
dim(mydata2) # qui ne perdiamo molti
mydata2<-mydata2[-which(!complete.cases(mydata2$lrscale)),]
dim(mydata2)
levels(mydata2$lrscale)<-c("left", "left", "left", "center-left", "center-left", "center", "center-
right", "center-right", "right", "right", "right", "Refusal", "Don't know", "No answer"
)
mydata2<-mydata2[-which(!complete.cases(mydata2$nwspol)),]
dim(mydata2)

```

```

mydata2<-mydata2[-which(!complete.cases(mydata2$trstplt)),]
dim(mydata2)
# trasformo in numerica
levels(mydata2$trstplt)<-c(0,1,2,3,4,5,6,7,8,9,10,"Refusal", "Don't know", "No answer"
)
mydata2$trstplt<-as.numeric(as.character(mydata2$trstplt))
mydata2$voto[which(mydata2$voto=="GUE/NGL")<-"NGL"
unique(mydata2$voto)
levels(mydata2$euftf)<-c(0,1,2,3,4,5,6,7,8,9,10,"Refusal", "Don't know", "No answer"
)
mydata2$euftf<-as.numeric(as.character(mydata2$euftf))
levels(mydata2$imwbcnt)<-c(0,1,2,3,4,5,6,7,8,9,10,"Refusal", "Don't know", "No an-
swer" )
mydata2$euftf<-as.numeric(as.character(mydata2$imwbcnt))
mydata2<-mydata2[-which(!complete.cases(mydata2$imwbcnt)),]
dim(mydata2)
# MCMCglmm estimates
# set seed
set.seed(123)
# def prior
# k<-length(levels(as.factor(data2$idchoice)))
k<-length(unique(mydata2$voto))
I <- diag(k-1)
J <- matrix(rep(1,(k - 1)2), c(k-1,k-1))
prior1<-list(R = list(fix=1, V=(1/6000) * (I + J)))
# A Modelli senza effetti random
# modello 1 no random
modell1<-MCMCglmm(voto~1+trait+agea+gndr, rcov= us(trait):units, data=mydata2,
family="categorical", prior=prior1, burnin=1000, nitt=11000)

```

```

summary(model1)
# modello 2 no random
model2<-MCMCglmm(voto -1+trait+agea+gndr+eisced, rcov= us(trait):units, da-
ta=mydata2, family="categorical", prior=prior1, burnin=1000, nitt=11000)
summary(model2)
# modello 3 no random
model3<-MCMCglmm(voto -1+trait+agea+gndr+eisced+hinctnta, rcov= us(trait):units,
data=mydata2, family="categorical", prior=prior1, burnin=1000, nitt=11000)
summary(model3)
# modello 4 no random, scala su posizionamento politico, tempo dedicato, fiducia,
immigrazione, integrazione
model4<-MCMCglmm(voto -1+trait+agea+gndr+eisced+hinctnta+lrscale + nwspol+trstplt
+euftf+imwbcnt, rcov= us(trait):units, data=mydata2, family="categorical", prior=prior1,
burnin=1000, nitt=11000)
summary(model4)
# B Modelli a effetti random
# definizione della prior
prior2<-list(R = list(fix = 1, V = (I+J),n = (k-1)),#(1/6000) * G=list(G1=list(V=diag((k-
1)), n=(k-1))))
# random effects cluster ward
set.seed(123)
# modello 1
model1r<-MCMCglmm(voto -1+trait+agea+gndr, rcov= us(trait):units, random= us(trait):
group, data=mydata2, family="categorical", prior=prior2, burnin=3000, nitt=33000)
summary(model1r)
#modello 2
set.seed(123)
model2r<-MCMCglmm(voto -1+trait+agea+gndr+eisced, rcov= us(trait):units, ran-
dom= us(trait):group, data=mydata2, family="categorical", prior=prior2, burnin=3000,
nitt=33000) summary(model2r)

```

```

# modello 3
set.seed(123)
model3r<-MCMCglmm(voto -1+trait+agea+gndr+eisced+hinctnta, rcov= us(trait):units,
random= us(trait):group, data=mydata2, family="categorical", prior=prior2, burnin=3000,
nitt=33000)
summary(model3r)
# modello 4 scala su posizionamento politico, tempo dedicato, fiducia, immigrazione,
integrazione
set.seed(123)
model4r<-MCMCglmm(voto -1+trait+agea+gndr+eisced+hinctnta+lrscale+nwspol+trstplt
+euftf+imwbent, rcov= us(trait):units, random= us(trait):group, data=mydata2, fa-
mily="categorical", prior=prior2, burnin=5000 nitt=35000, thin=15)
summary(model4r).

```

C.2 Tabelle excel di supporto al codice

La prima tabella riportata, associa i vari partiti nazionali alle rispettive famiglie europee, mentre la seconda evidenzia le famiglie europee rappresentate in ogni Paese.

Bulgaria	Gruppo Europeo	Francia	Gruppo Europeo
Grazhdani za evropeitsko razvitie na Bulgariya (GERB)"	Altro	LO (Lutte Ouvri�re)	GUE/NGL
Balgarska sotsialisticheska partiya (BSP)	S&D	NPA (Nouveau Parti Anti-Capitaliste)	GUE/NGL
Obedineni patrioti - NFSB, Ataka i VMRO	ECR	PCF (PARTI COMMUNISTE FRAN�AIS)	GUE/NGL
Dvizhenie za prava i svobodi (DPS)	ADLE	FI (LA FRANCE INSOUmise)	GUE/NGL
Volya	Altro	PS (Parti Socialiste)	GUE/NGL
Reformatorski blok - Glas Naroden	PPE	EELV (Europe Ecologie Les Verts)	Verdi/ALE
Koalitsiya Dvizhenie za Bulgariya	S&D	LREM (LA R�PUBLIQUE EN MARCHE)	ADLE
Obedinenie DOST	Altro	MODEM (Mouvement D�mocrate)	ADLE
Nova republika - DSB - Suyuz za Plovdiv - BDO	PPE	LR (LES R�PUBLICAINS)	PPE
Koalitsiya ABV �c" Dvizhenie 21	Altro	DEBOUT LA FRANCE	EFDD
Vuzrazhdane	Altro	FN (Front National)	ENL
Italia	Gruppo Europeo	Svizzera	Gruppo Europeo
Partito Democratico (PD)	S&D	Swiss People's Party	Altro
+ Europa	ADLE	Social Democratic Party / Socialist Party	S&D
Italia Europa Insieme		FDP. The Liberals	ADLE
Civita Popolare Lorenzin	PPE	Christian Democratic Party	PPE
SVP-PATT	PPE	Green Party	Verdi/ALE
Liberi e Uguali	GUE/NGL	Green Liberal Party	ADLE
Movimento 5 Stelle	EFDD	Conservative Democratic Party	Altro
Forza Italia	PPE	Evangelical People's Party	ECR
Lega Nord	ENL	Federal Democratic Union	ECR
Fratelli d'Italia	ECR	Ticino League	Altro
Noi con l'Italia - UDC	PPE	Swiss Labour Party	GUE/NGL
Potere al popolo	GUE/NGL	Movement of the Citizens of French-speaking Switzerland	Altro
Casapound Italia	ECR	Alternative Left	GUE/NGL
		Pirate Party Switzerland	Verdi/ALE
Austria	Gruppo Europeo	Cipro	Gruppo Europeo
SP�-	S&D	Progressive Party of Working People (AKEL)	GUE/NGL
�-VP	PPE	Democratic Party (DIKO)	S&D
FP�-	ENL	Democratic Rally (DISY)	PPE
PILZ	Verdi/ALE	National Popular Front (ELAM)	Altro
Gr�ne	Verdi/ALE	Solidarity Movement (KINIMA ALLILEGII)	ECR
KP�-	GUE/NGL	The Cyprus Green Party	Verdi/ALE
NEOS	ADLE	Citizens' Alliance (SYMMACHIA POLITON)	S&D
GILT	Altro	Social Democrats (KS EDEK)	S&D

Belgio	Gruppo Europeo	Portogallo	Gruppo Europeo
Groen!	Verdi/ALE	PTP-MAS - Agir	Altro
CD&V	PPE	B.E. - Bloco de Esquerda	GUE/NGL
N-VA	ECR	PCP-PEV - CDU - ColigaÃ§Ã£o DemocrÃ¡tica UnitÃ¡ria	GUE/NGL
Lijst Dedecker	ECR	JPP - Juntos pelo Povo	Altro
SP.A	S&D	L/TDA - LIVRE/Tempo de AvanÃ§ar	Altro
PVDA	GUE/NGL	NC - NÃ¡s, CidadÃ£os!	S&D
Vlaams Belang	ENL	PPV/CDC - Partido Cidadania e Democracia CristÃ£	Altro
Open VLD	ADLE	PCTP/MRPP - Partido Comunista dos Trabalhadores Portugueses	Altro
CDH	PPE	MPT - Partido da Terra	PPE
Ecolo	Verdi/ALE	PDR - Partido DemocrÃ¡tico Republicano	ADLE
Front National	Altro	PNR - Partido Nacional Renovador	Altro
MR	ADLE	PPM - Partido Popular MonÃ¡rquico	Altro
PS	S&D	PS - Partido Socialista	S&D
PTB	GUE/NGL	PURP - Partido Unido dos Reformados e Pensionistas	Altro
Parti Populaire	Altro	PAN - Pessoas-Animais-Natureza	Verdi/ALE
DÃ©FI	Altro	PPD-PSD/CDS-PP - Portugal Ã Frente	PPE
Slovacchia	Gruppo Europeo	Germania	Gruppo Europeo
ObyÄdnÄ Ã¼dia a nezÄvislÄ osobnosti	ECR	Christian Democratic Union/Christian Social Union (CDU/CSU)	PPE
SlovenskÃ¡ nÃ¡rodnÃ¡ strana (SNS)	Altro	Social Democratic Party (SPD)	S&D
Smer Ä SD	S&D	The Left (Die Linke)	GUE/NGL
MostÄ HÃ¡	Altro	Alliance 90/The Greens (BÃ¼ndnis 90/Die GrÃ¼nen)	Verdi/ALE
ÄS NaÄe Slovensko	Altro	Free Democratic Party (FDP)	ADLE
Sloboda a Solidarita (SaS)	ECR	Alternative for Germany (AFD)	EFDD
SieÄy	Altro	Pirate Party (Piratenpartei)	Verdi/ALE
SME Rodina	Altro	National Democratic Party (NPD)	Altro
Estonia	Gruppo Europeo	Norvegia	Gruppo Europeo
Eesti Reformierakond	ADLE	RÃ¡dt	GUE/NGL
Eesti Keskerakond	ADLE	Sosialistisk Venstreparti	Verdi/ALE
Erakond Isamaa ja Res Publica Liit	PPE	Arbeiderpartiet	S&D
Sotsiaaldemokraatlik Erakond	S&D	Venstre	ADLE
Erakond Eestimaa Rohelised	Verdi/ALE	Kristelig Folkeparti	PPE
Eesti Konservatiivne Rahvaerakond	ENL	Senterpartiet	ADLE
Eesti Iseseisvusparte	Altro	HÃ¡yre	PPE
Äeksikkandidaadid vÃ¡pi muud	Altro	Fremskrittspartiet	Altro
Eesti Vabaerakond	ECR	Kystpartiet	Altro
Rahva Ähtsuse Erakond		MiljÃ¡partiet De GrÃ¼nne	Verdi/ALE

Spagna	Gruppo Europeo	UK	Gruppo Europeo
PP	PPE	Conservative	PPE
PSOE	S&D	Labour	S&D
Unidas Podemos	GUE/NGL	Liberal Democrat	ADLE
Podemos	GUE/NGL	Scottish National Party	Verdi/ALE
Ciudadanos	ADLE	Plaid Cymru	Verdi/ALE
IU	GUE/NGL	Green Party	Verdi/ALE
En Com� Podem	Altro	UK Independence Party	ENL
M�s Pa�s	Altro	Ulster Unionist Party (nir)	ECR
Comprom�s	Verdi/ALE	Democratic Unionist Party (nir)	Altro
ERC	Verdi/ALE	Sinn F�in (nir)	GUE/NGL
Junts per Catalunya/PDeCAT	ADLE	Social Democratic and Labour Party (nir)	EFDD
Bloque Nacionalista Galego	Verdi/ALE	Alliance Party (nir)	ADLE
EAJ-PNV	ADLE	Traditional Unionist (nir)	S&D
EH-Bildu	GUE/NGL	Green Party (nir)	Verdi/ALE
CC-PNC	S&D	Independent(s) (nir)	Altro
VOX	ECR	People Before Profit Alliance (nir)	GUE/NGL
Navarra Suma	Altro	Progressive Unionist Party	Altro
PACMA	Verdi/ALE		
CUP	Altro		
PRC	GUE/NGL		
Finlandia	Gruppo Europeo	Olanda	Gruppo Europeo
The National Coalition Party	PPE	People's Party for Freedom and Democracy	ADLE
The Swedish People's Party (SPP)	ADLE	Labour Part	S&D
The Centre Party	ADLE	Party for Freedom	ENL
True Finns	ECR	Socialist Party	GUE/NGL
Christian Democrats	PPE	Christian Democratic Appeal	PPE
Freedom Party	Altro	Democrats '66	ADLE
Pirate Party	Verdi/ALE	Christian Union	ECR
Independence Party	Altro	Green Left	Verdi/ALE
For the Poor	Altro	Reformed Political Party	ECR
Green League	Verdi/ALE	Party for the Animals	GUE/NGL
Social Democratic Party	S&D	50PLUS	Altro
Left Alliance	GUE/NGL	DENK	Altro
Communist Party	Altro	Forum for Democracy	ECR
The Communist Workers' Party	Altro	Article 1	Altro
Workers' Party	Altro		

Croazia	Gruppo Europeo	Irlanda	Gruppo Europeo
Hrvatska demokratska zajednica (HDZ)	PPE	Anti-Austerity Alliance - People Before Profit	GUE/NGL
Narodna koalicija - SDP, HNS, HSU, HSS	S&D	Fianna Fáil	ADLE
Most nezavisnih lista (Most)	ECR	Fine Gael	PPE
Koalicija "Mivi zid, Promijenimo Hrvatsku, Abeceda, HDSS	Altro	Green Party	Verdi/ALE
Koalicija za premijera - BM 365, NS-R, NV-SR, HSS-SR, BUZ	Altro	Independent	Altro
Koalicija IDS, PGS, LRI	Altro	Labour	S&D
Koalicija HDSSB-HKS	Altro	Sinn Féin	GUE/NGL
Samostalna demokratska srpska stranka (SDSS)	S&D	Social Democrats	S&D
Koalicija Pametno i Za grad	ADLE	Socialist Party - United Left Alliance	GUE/NGL
Ungheria	Gruppo Europeo	Slovenia	Gruppo Europeo
DK (Demokratikus Koalíció)	S&D	DESUS - Demokratska stranka upokojencev Slovenije	ADLE
Együtt2014 Mozgalom	Verdi/ALE	L - Levica	GUE/NGL
Fidesz (Fidesz Magyar Polgári Párt)	PPE	LM - Lista Marjana Arca	ADLE
Jobbik (Jobbik Magyarországiért Mozgalom)	Altro	NSI - Nova Slovenija "Kršćanski demokrati"	PPE
KDNP (Kereszténydemokrata Néppárt)	PPE	PS - Pozitivna Slovenija	Altro
LMP (Lehet Más A Politika)	Verdi/ALE	SD - Socialni demokrati	S&D
SZP (Magyar Szocialista Párt)	S&D	SAB - Stranka Alenke Bratušek	ADLE
Munkáspárt (Magyar Kommunista Munkáspárt)	Altro	SDS - Slovenska demokratska stranka	PPE
Momentum Mozgalom	ADLE	SLS - Slovenska ljudska stranka	PPE
Párbeszéd (Párbeszéd Magyarországiért Párt)	Verdi/ALE	SMC - Stranka modernega centra	ADLE
MKKP (Magyar Kötőfarkú Kutya Párt)	Altro	SNS - Slovenska nacionalna stranka	Altro

Polonia	Gruppo Europeo	Svezia	Gruppo Europeo
KORWiN	ECR	Centern	ADLE
Kukiz'15	Altro	Folkpartiet liberalerna	ADLE
Nowoczesn	Altro	Kristdemokraterna	PPE
Platforma Obywatelska	PPE	MiljÅpartiet de grÅna	Verdi/ALE
Polskie Stronnictwo Ludowe	PPE	Moderata samlingspartiet	PPE
Prawo i SprawiedliwoÅ	ECR	Socialdemokraterna	S&D
Razem	S&D	VÅnsterpartiet	GUE/NGL
Zjednoczona Lewica	Altro	FI (Feministiskt initiativ)	S&D
		Sverigedemokraterna	ECR
		Annat parti	Altro
Serbia	Gruppo Europeo		
Aleksandar VuÅ\u008diÅ+- Srbija pobeÅ'uje	Altro		
Boris Tadic, Cedomir Jovanovic - Savez za bolju Srbiju - LDP, LSV, SDS	ADLE		
Borko StefanoviÅ+ - Srbija za sve nas	Altro		
Dosta je bilo - SaÅ;a RaduloviÅ+	ECR		
Dr Vojislav Å eÅ;elj - Srpska radikalna stranka	Altro		
Dveri - Demokratska stranka Srbije - Sanda RaÅ;koviÅ+ IviÅ+ - BoÅ;ko ObradoviÅ+	Altro		
Ivica DaÅ\u008diÅ+ - 'SocijalistiÅ\u008dka partija Srbije (SPS), Jedinstvena Srbija (JS)-	Altro		
Muamer ZukorliÅ+ - BoÅ;njaÅ\u008dka demokratska zajednica SandÅ;kak	Altro		
Partija za demokratsko delovanje - Ardita Sinani/ Partia pÅkr vertim demokratik	Altro		
SDA SandÅ;kaka - Dr. Sulejman Ugljanin	Altro		
VajdasÅ;gi Magyar SzÅtvetsÅg-PÅ;sztor IstvÅ;n - Savez vojvoÅ'anskih MaÅ'ara - I	PPE		
Za pravednu Srbiju - Demokratska stranka (NOVA, DSHV, ZS)	S&D		

Figura C.1: Associazione tra partiti nazionali e gruppi europei.

NUMERO	PAESE	LIVELLI	NUMERO	PAESE	LIVELLI	NUMERO	PAESE	LIVELLI
1	ITALIA	S&D	5	PAESI BASSI	ADLE	9	ESTONIA	ADLE
		ADLE			S&D			PPE
		PPE			ENL			S&D
		GUE/NGL			GUE/NGL			VERDI/ALE
		EFDD			S&D			ENL
		ENL			ADLE			PPE
		ECR			PPE			S&D
2	FRANCIA	GUE/NGL	6	SVIZZERA	VERDI/ALE	10	GRAN BRETAGNA	ADLE
		VERDI/ALE			ECR			VERDI/ALE
		ADLE			GUE/NGL			ENL
		PPE			PPE			ECR
		EFDD			S&D			GUE/NGL
		ENL			GUE/NGL			EFDD
					ADLE			S&D
3	BULGARIA	PPE	6	SPAGNA	VERDI/ALE	11	UNGHERIA	VERDI/ALE
		S&D			ECR			PPE
		ECR			ADLE			ADLE
		ADLE			PPE			ADLE
	SERBIA	ADLE		FINLANDIA	ADLE	12	IRLANDA	GUE/NGL
		ECR			ECR			ADLE
		PPE			VERDI/ALE			PPE
		S&D			S&D			VERDI/ALE
		PPE			GUE/NGL			S&D
		S&D			GUE/NGL			S&D
CROAZIA	ECR	SVEZIA	ADLE	12	NORVEGIA	GUE/NGL		
	ADLE		PPE			VERDI/ALE		
			VERDI/ALE			S&D		
			S&D			ADLE		
4	AUSTRIA	ECR	7	CIPRO	GUE/NGL	13	POLONIA	PPE
		PPE			ECR			ECR
		S&D			GUE/NGL			PPE
		VERDI/ALE			S&D			S&D
		GUE/NGL			PPE			GUE/NGL
5	BELGIO	ADLE	8	GERMANIA	VERDI/ALE	14	PORTOGALLO	S&D
					PPE			PPE
					S&D			ADLE
					GUE/NGL			VERDI/ALE
					GUE/NGL	15	SLOVENIA	ADLE
		ENL			VERDI/ALE			GUE/NGL
					ADLE			PPE
					ADLE			S&D
		EFDD	SLOVACCHIA	ECR				
				S&D				

Figura C.2: Paesi e gruppi europei.