

UNIVERSITÀ DEGLI STUDI DI MILANO



Metodi e modelli per le decisioni

Roberto Cordone

Indice

I Problemi decisionali	13
1 Introduzione	15
1.1 Terminologia	15
1.2 L'approccio modellistico alla decisione	16
1.3 Perché un approccio formale?	17
1.4 Modelli prescrittivi e modelli descrittivi	18
1.5 Business Intelligence	19
1.6 Esempi di problemi decisionali complicati	21
1.7 Classificazione dei problemi decisionali	22
2 Casi di studio	25
2.1 La Metrotranvia di Como	26
2.1.1 Il contesto	26
2.1.2 La generazione delle alternative	31
2.1.3 La generazione degli scenari	36
2.1.4 La definizione e il calcolo degli indicatori	38
2.1.5 La definizione degli attori	40
2.1.6 Il processo di decisione	45
2.2 La riapertura dei Navigli a Milano	46
2.2.1 Il contesto	46
2.2.2 La definizione delle alternative	48
2.2.3 La definizione degli scenari	50
2.2.4 La definizione e il calcolo degli indicatori	50
2.2.5 L'organizzazione spaziale e temporale	51
2.2.6 La definizione degli attori	52
2.2.7 Il processo di decisione	52
3 Definizioni fondamentali e problemi concettuali	53
3.1 Alternative	53
3.2 Scenari	54
3.3 Impatti e funzione impatto	55
3.4 Decisori	57
3.5 Relazione di preferenza	57
3.5.1 Rappresentazioni della preferenza nel caso finito	58

3.5.2	Relazioni derivate	58
3.5.3	Proprietà della relazione di preferenza	60
3.5.4	Relazioni di preferenza più comuni	61
3.5.5	Problemi critici sull'ipotesi di transitività	62
3.5.6	Ordini deboli e funzioni valore	64
3.6	Dominanza fra soluzioni	69
3.7	Esercizi	71
4	Programmazione Matematica	73
4.1	Concetti di base	74
4.1.1	Sviluppo in serie di Taylor	75
4.1.2	Direzioni	76
4.2	Condizioni necessarie di ottimalità locale	76
4.2.1	Condizione di non miglioramento	77
4.2.2	Condizione di ammissibilità	79
4.2.3	Una prima interpretazione geometrica	85
4.3	Condizioni di Karush-Kuhn-Tucker	86
4.3.1	Lemma di Farkas	86
4.3.2	Forma standard delle condizioni di Karush-Kuhn-Tucker	87
4.3.3	Una seconda interpretazione geometrica	89
4.4	Applicazioni delle condizioni di KKT	90
4.4.1	Pianificazione per sistemi Big-Data	90
4.5	Esercizi	94
II	Modelli con preferenze complesse	117
5	La preferenza paretiana	119
5.1	Proprietà formali della preferenza paretiana	119
5.2	Dominanza paretiana	120
5.3	Determinazione della regione paretiana	121
5.3.1	L'applicazione della definizione	121
5.4	Metodo della trasformazione inversa	122
5.5	Le condizioni di Karush-Kuhn-Tucker	124
5.6	Metodo dei pesi	134
5.6.1	Vantaggi e svantaggi	137
5.6.2	Il metodo dei pesi nella Programmazione Lineare	138
5.6.3	Analisi di sensitività	140
5.7	Metodo dei vincoli	141
5.7.1	Vantaggi e svantaggi	144
5.7.2	Il metodo dei vincoli nella Programmazione Lineare	145
5.8	Esercizi	148
6	Strutturazione delle preferenze	155
6.1	Ordine lessicografico	155

6.2	Ordine lessicografico con livelli di aspirazione	156
6.3	Punto utopia	158
6.4	Teoria dell'utilità a molti attributi	159
6.4.1	Curve di indifferenza	160
6.4.2	Determinare la funzione di utilità	161
6.4.3	Funzioni di utilità additive	162
6.4.4	Indipendenza preferenziale	163
6.4.5	Indipendenza preferenziale e additività	165
6.4.6	Tasso di sostituzione	167
6.4.7	Additività e tasso di sostituzione	168
6.4.8	Costruzione di una funzione di utilità additiva	171
6.4.9	Il procedimento in sintesi	176
6.5	Esercizi	178
7	Metodi a razionalità debole	185
7.1	Decisori parzialmente incoerenti	185
7.1.1	La ricostruzione di matrici coerenti	188
7.1.2	Coerenza ad ogni costo?	190
7.2	Analisi Gerarchica	190
7.2.1	Determinazione delle utilità tramite confronti a coppie	191
7.2.2	Scale qualitative	191
7.2.3	Determinazione dei pesi con confronti a coppie	192
7.2.4	Strutturazione gerarchica degli attributi	192
7.2.5	Ricomposizione gerarchica	193
7.2.6	Ibridazioni con la teoria dell'utilità	195
7.2.7	Rank reversal	201
7.2.8	Scale assolute o metodo della stima a priori	202
7.2.9	Rank reversal nella teoria dell'utilità	204
7.3	Metodi ELECTRE	205
7.3.1	La relazione di surclassamento	206
7.3.2	Raffinamento della relazione di surclassamento	207
7.3.3	Costruzione del nucleo	210
7.3.4	Creazione di un ordinamento debole	212
7.4	Esercizi	214
III	Modelli con scenari molteplici	219
8	Modellazione dell'incertezza	221
8.1	Modellazione dell'incertezza	222
8.2	Dominanza forte	224
8.3	Exercises	226
9	Decisioni in condizioni di ignoranza	227
9.1	Criterio del caso pessimo	228

9.2 Criterio del caso ottimo	228
9.3 Criterio di Hurwicz	228
9.3.1 Taratura del coefficiente di pessimismo	229
9.3.2 Sensitività rispetto al coefficiente di pessimismo	230
9.4 Criterio di equiprobabilità	231
9.5 Criterio del rammarico	232
9.6 Criterio delle eccedenze	233
9.7 Esempi	234
9.7.1 Installazione di un dispositivo	234
9.7.2 Un investimento finanziario	238
9.7.3 Un esempio continuo	241
9.8 Difetti formali delle logiche di scelta	246
9.8.1 Dipendenza da alternative irrilevanti	246
9.8.2 Dipendenza da duplicazioni di scenari	248
9.8.3 Dipendenza da variazioni uniformi in uno scenario	249
9.9 La Programmazione Robusta	250
9.9.1 Criteri adottati in Ottimizzazione Combinatoria	250
9.9.2 Risultati teorici di complessità	252
10 Programmazione in condizioni di rischio	261
10.1 Criterio del valore atteso	261
10.2 Sensitività rispetto alle probabilità	262
10.3 Difetti formali del criterio del valore atteso	264
10.3.1 Incoerenza fra valore atteso ed equivalente certo	264
10.3.2 La scommessa di Pascal	265
10.3.3 Il paradosso di San Pietroburgo	266
10.4 Teoria dell'utilità stocastica	267
10.4.1 Assiomi fondativi dell'utilità stocastica	268
10.4.2 Il teorema dell'utilità stocastica di Von Neumann e Morgenstern	269
10.5 Avversione e propensione al rischio	272
10.5.1 Equivalente certo e premio di rischio	273
10.6 Critiche alla teoria dell'utilità stocastica	274
10.6.1 Paradosso di Allais	275
10.6.2 Il paradosso di Ellsberg	276
11 Teoria delle decisioni	279
11.1 Albero delle decisioni	279
11.2 Scenari condizionati dalla decisione	284
11.3 Decisioni distribuite in più fasi	284
11.4 Esperimenti casuali	286
11.4.1 Calcolo delle probabilità per l'albero delle decisioni	288
11.5 Exercises	293

IV Modelli con decisori molteplici	303
12 Teoria dei giochi	305
12.1 Giochi in forma estesa	307
12.2 Giochi in forma strategica	310
12.2.1 Dominanza fra strategie	312
12.2.2 Equilibrio	315
12.3 La strategia del caso pessimo	316
12.4 Giochi di interesse particolare	318
13 Giochi a somma zero	319
13.1 Strategie dominate	320
13.2 Equilibri	320
13.3 Valore del gioco	320
13.4 Strategie miste	322
13.5 Teorema del minimax di Von Neumann	324
13.5.1 Determinazione della strategia mista di equilibrio	327
13.5.2 Metodo del pivot	328
14 Giochi simmetrici	333
14.1 Tassonomia dei giochi simmetrici a due persone e due strategie	334
14.1.1 Classificazione in base agli equilibri di Nash	334
14.1.2 Classificazione in base all'ordinamento dei payoff	335
14.2 Il matrimonio perfetto	335
14.3 La caccia al cervo	336
14.4 I giochi di coordinamento puro	337
14.5 La corsa del coniglio	338
14.6 La guerra dei sessi	339
14.7 Il dilemma del prigioniero	340
14.8 Altri giochi	341
14.8.1 Il gioco dei maiali nel recinto	341
14.9 Giochi finiti e strategie miste	342
14.10 Exercises	343
15 Teoria delle decisioni di gruppo	353
15.1 Costituzione	354
15.2 Metodo di Condorcet	355
15.3 Metodo di Borda	356
15.4 Metodo lessicografico	358
15.5 Sistema di pluralità	358
15.6 L'approccio assiomatico	360
15.7 Critiche agli assiomi di Arrow	364
15.7.1 Critiche all'assioma di non banalità	365
15.7.2 Critiche all'assioma del dominio universale	365
15.7.3 Critiche all'assioma della rilevanza binaria	366

15.7.4 Critiche all'assioma di unanimità	367
15.8 Impossibilità per i sistemi elettorali	367
V Modelli descrittivi	369
11 Modelli per i sistemi di trasporto	371
11.1 Modello di generazione	373
11.2 Modello di distribuzione	374
11.3 Modello di scelta modale	376
11.4 Modello di scelta del percorso	377
12 Modelli di teoria delle code	383
13 Modelli di simulazione a eventi discreti	385
14 Modelli dinamici	387
A Richiami di Analisi Matematica	389
A.1 Richiami sulle funzioni di una sola variabile	389
A.2 Funzioni di più variabili	390
A.3 Linee in forma parametrica	394
B Richiami di Algebra Lineare	397
B.1 Combinazioni di vettori	397
C Richiami di calcolo delle probabilità	399

Prefazione

L'etichetta *Metodi e modelli per le decisioni* è applicata nelle università italiane a corsi piuttosto vari nel contenuto. Spesso si tratta di un altro nome per corsi di Ricerca Operativa, cioè di modelli e algoritmi per l'ottimizzazione, oppure di corsi che presentano raccolte di applicazioni pratiche della Ricerca Operativa. Questo corso si muove nello stesso ambito generale, ma con una fondamentale differenza.

Il tema comune di questi corsi sta nel concetto di *decisione*, cioè di scelta fra alternative non indifferenti fra loro. È una situazione che tutti viviamo continuamente, ma che in alcuni casi risulta particolarmente difficile, perché i dati da considerare nella scelta sono molti, perché le alternative possibili sono molte e perché i costi di una scelta errata sono alti. In questi casi, la strategia corretta non è affidarsi al buon senso, all'esperienza passata o al caso, ma modellare il problema, calcolare la soluzione, interpretarla e solo al termine decidere. Questo si sa da decenni (almeno dalla seconda guerra mondiale). Oggi però nuovi strumenti hanno riportato di moda l'approccio modellistico alla decisione: il *Big Data* offre enormi quantità di dati precisi, strutturati e poco costosi, che aspettano solo di essere sfruttati per estrarne informazione; il *Cloud Computing* offre accesso ai dati e potenza di calcolo distribuiti ovunque a basso costo; la *Business Analytics* propaganda il vantaggio competitivo di una cultura aziendale aperta all'uso dei modelli per ricavare sulla propria azienda, sui clienti, sui concorrenti e sull'ambiente in cui si opera tutte le informazioni utili a definire le proprie strategie operative. Tutto questo ha stimolato nell'accademia lo sviluppo di nuovi filoni di ricerca teorica: l'ottimizzazione online, l'ottimizzazione stocastica, l'ottimizzazione robusta, l'ottimizzazione multilivello, ecc...

La differenza fondamentale tra questo corso e la maggior parte di quelli omonimi è che non si occupa degli aspetti tecnici di risoluzione dei modelli decisionali, cioè delle proprietà matematiche e degli algoritmi. Discute invece:

- i fattori che complicano in linea di principio la costruzione di tali modelli;
- i metodi matematici proposti per trattare tali complicazioni e i modelli matematici che ne derivano;
- i limiti e gli errori intrinseci di tali metodi e modelli, compresi alcuni veri e propri risultati di impossibilità a costruire soluzioni soddisfacenti.

Questo corso cioè intende mostrare perché usare modelli per aiutarsi nelle decisioni è utile e perché i risultati che producono vanno letti con spirito critico. Si sposta l'accento dal nucleo duro del problema decisionale a tutto ciò che circonda il problema e che in molti casi è indispensabile prendere in considerazione per poter arrivare a tale nucleo e per poter interpretare in modo sensato i risultati degli algoritmi applicati ad esso.

Questo corso ha due sorgenti fondamentali. La prima è il corso di *Metodi e modelli per il supporto alle decisioni* (dove “supporto” è un calco dell'inglese “support”

e sta per l'italiano “ausilio”) che si teneva presso la sede di Como del Politecnico di Milano, a beneficio di studenti in ingegneria ambientale e informatica e per il quale anni fa tenni qualche attività secondaria. Da questa sorgente deriva la maggior parte degli argomenti e la scelta dei casi di studio. Questi ultimi riguardano le decisioni nell'ambito delle grandi opere pubbliche, che sono probabilmente le situazioni in cui è più chiara e ineludibile la difficoltà di passare dal problema concreto a un modello di ottimizzazione pura e semplice e dai risultati numerici del modello a una strategia operativa. D'altra parte, questi casi non esauriscono i campi applicativi nei quali i modelli decisionali fanno da guida per l'azione. Altri esempi sono:

- in finanza, dove si usano per definire strategie di investimento;
- nel marketing, dove si usano per definire campagne pubblicitarie o strategie di pricing;
- nella gestione delle risorse naturali, dove si usano per definire strategie di prelievo;
- nei videogiochi, dove si usano per definire strategie di gioco.

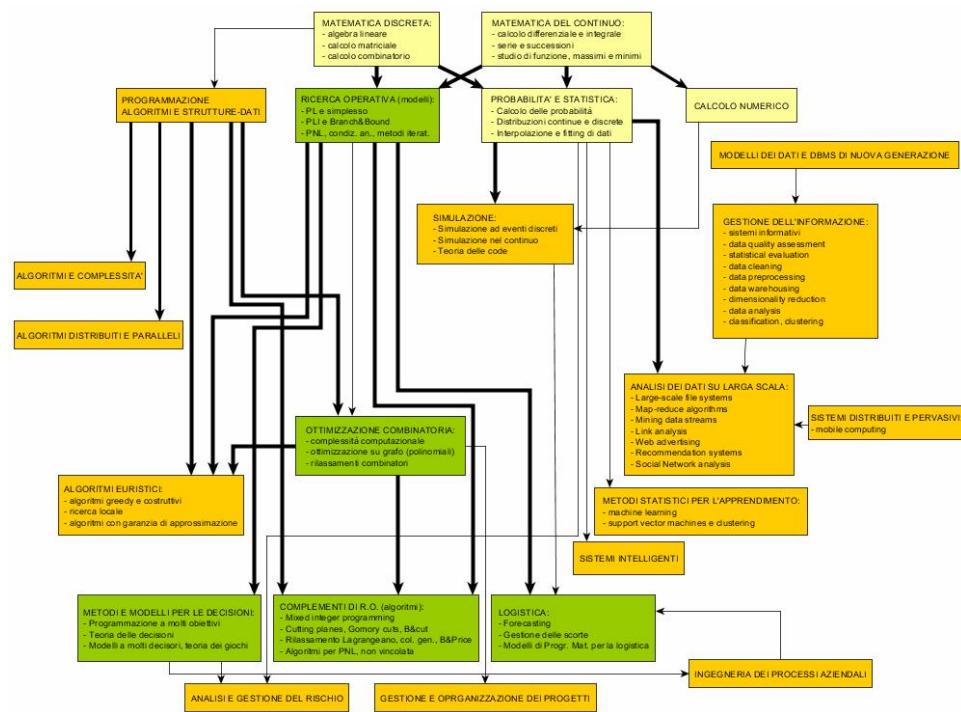


Figura 1: Struttura dell'indirizzo in *Analytics e ottimizzazione*

La seconda sorgente del corso è l'indirizzo in *Analytics e ottimizzazione* che i corsi di laurea in Informatica e Matematica offrono da alcuni anni e nel quale questo corso si inquadra con un ruolo ben determinato. Da questa sorgente derivano alcuni argomenti, ma soprattutto un'impostazione più rigorosa e formale e relazioni di complementarietà con gli altri corsi dell'indirizzo. Anticipo che il corso è strutturato in modo da poter essere seguito e risultare interessante e utile anche per studenti di altri indirizzi e di altri corsi di laurea, ma dedico un paragrafo all'indirizzo di *Analytics e ottimizzazione* per chiarire meglio il taglio e lo scopo fondamentale di

questo corso. La Figura 1 illustra la struttura dell'indirizzo: i corsi in giallo indicano materie di base, quelli in arancione materie informatiche affini¹, quelli in verde (più *Algoritmi euristici*) sono i corsi fondamentali dell'indirizzo. Schematicamente:

- *Ricerca operativa* illustra il concetto di modello decisionale e l'uso dei risolutori generici;
- *Ottimizzazione combinatoria* si concentra sui problemi per i quali è stata dimostrata l'esistenza di algoritmi polinomiali e discute gli algoritmi e le strutture dati migliori note per i problemi fondamentali;
- *Complementi di ricerca operativa* si concentra sui problemi per i quali non si conoscono algoritmi polinomiali e discute il progetto di algoritmi esponenziali che nel caso medio sono in grado di risolvere istanze di dimensione non grande, ma utile in pratica;
- *Algoritmi euristici* si concentra sui problemi per i quali non si conoscono algoritmi polinomiali e discute il progetto di algoritmi polinomiali che non garantiscono la soluzione ottima, ma nel caso medio calcolano soluzioni di buona qualità;
- *Logistica* presenta una raccolta di modelli e algoritmi per uno specifico ambito applicativo, cioè la produzione, l'immagazzinamento e il trasporto delle merci;
- *Metodi e modelli per le decisioni* fa da corona a tutto questo, discutendo come costruire i modelli decisionali discussi negli altri corsi e fino a che punto tali modelli rappresentino correttamente la realtà.

Da questo schema dovrebbe risultare chiaro che il corso di *Metodi e modelli per le decisioni* non richiede forti competenze sulle proprietà dei modelli e degli algoritmi di ottimizzazione. Gli studenti che non le avessero si limiteranno ad accettare la loro esistenza come scatole nere. Gli studenti che invece ne sapessero di più, per esempio per aver seguito gli altri corsi dell'indirizzo, useranno questo corso per allargare la loro visione a ciò che precede e segue la fase di risoluzione algoritmica. Queste dispense includono diverse sezioni contrassegnate da un asterisco che vanno intese come approfondimenti e non sono richieste ai fini dell'esame: tali sezioni offrono legami interessanti con gli altri corsi dell'indirizzo.

A testimonianza della interdisciplinarietà del corso e della possibilità di seguirlo partendo da formazioni anche molto diverse, oltre ad essere aperto agli studenti di Informatica e di Matematica, a partire dal 2018 il corso è mutuato dalla laurea in *Data Science and Economics*, come modulo di *Optimization* nel corso di *Graph Theory, Discrete Mathematics and Optimization*. Tutto questo spinge necessariamente la trattazione più verso l'ampiezza che la profondità e implica che ogni gruppo di studenti trovi nel corso argomenti che già conosce o che gli paiono trattati in modo troppo elementare. Gli studenti di matematica troveranno i metodi analitici troppo informali, gli studenti di informatica troveranno appena accennati i temi del trattamento dei dati e della complessità computazionale, gli studenti di Data Science and Economics troveranno superficiale la trattazione economica. Questa conseguenza purtroppo è difficile da eliminare, se si vuole che tutti gli studenti abbiano la possibilità di seguire il corso e sostenere l'esame nei tempi stabiliti. D'altronde, è probabile che tutti trovino interessante la possibilità di ampliare la propria prospettiva attraverso la presentazione combinata dei diversi aspetti che in una decisione reale complicata devono necessariamente interagire.

¹Salvo *Algoritmi euristici* che è in arancione per motivi burocratici.

Parte I

Problemi decisionali

Capitolo 1

Introduzione

1.1 Terminologia

Informalmente, una decisione è una scelta fra alternative non indifferenti fra loro. Introduciamo un po' di termini che verranno usati nel seguito per essere più precisi e non ambigui. Non si tratta ancora di definizioni formali, ma solo dell'introduzione di alcune parole chiave: la Sezione 1.6 e il Capitolo 56 forniranno esempi concreti dell'uso di tali termini, mentre il Capitolo 3 ne darà la definizione formale e discuterà diversi problemi metodologici che il loro uso comporta. Purtroppo, i temi trattati nel corso sono studiati in molti settori di ricerca diversi (matematica, economia, ingegneria, ecc...) e ogni settore ha adottato una propria nomenclatura e proprie convenzioni, spesso diverse da quelle degli altri settori. Si è deciso di riportare tutti i termini di uso comune e di adottare in ogni capitolo quelli tipici del settore che ha più approfondito l'argomento specifico del capitolo stesso, richiamando ogni tanto l'equivalenza con altri termini. Questo viola la regola classica di usare una sola parola per ogni concetto, ma dovrebbe consentire agli studenti di accedere più facilmente alla letteratura. Alcuni concetti, inoltre, non hanno nomi standard, almeno a mia conoscenza, e mi sono quindi trovato a inventare una parola *ad hoc* per il corso.

Definiremo *sistema* la porzione del mondo sulla quale la decisione influisce¹. Un sistema non è dato una volta per tutte, altrimenti sarebbe possibile alcuna decisione. Indicheremo con *alternativa* o *soluzione* l'insieme degli aspetti controllabili del sistema², con *scenario* o *esito* l'insieme di quelli non controllabili³. Un sistema, quindi, combina aspetti controllabili e incontrollabili in una *configurazione*⁴. Ogni configurazione è caratterizzata da un *impatto*, che descrive tutti gli aspetti interessanti ai fini della decisione⁵. *Decisore* o *attore* è chiunque contribuisce alla scelta dell'alternativa: il primo termine indica chi partecipa alla scelta, mentre il secondo si allarga a includere anche chi non vi partecipa, ma ha degli interessi in gioco e può reagire, esercitando un'influenza indiretta sulla scelta stessa⁶. Infine, con *preferenza* intendiamo la descrizione della soddisfazione relativa fra impatti diversi⁷.

Un *problema decisionale* richiede di scegliere un'alternativa in modo da spingere il sistema in una configurazione il cui impatto sia preferito dai decisori rispetto a

¹ *Sistema* è termine tipico del lessico ingegneristico o matematico.

² *Alternativa* è più usato nel lessico economico, *soluzione* in quello matematico e ingegneristico.

³ *Scenario* è più usato nel lessico economico, *esito* in quello statistico.

⁴ *Configurazione* è un termine coniato *ad hoc*, benché sia a volte usato in ingegneria.

⁵ *Impatto* è un altro termine coniato *ad hoc*; in letteratura spesso si usano delle perifrasi.

⁶ Anziché *attore* si usa anche di frequente l'espressione inglese *stakeholder*.

⁷ *Preferenza* deriva dal lessico economico, ma è di uso comune nel linguaggio quotidiano.

quelli delle altre configurazioni, tenendo conto del fatto che la configurazione non dipende solo dall'alternativa scelta, ma anche dagli scenari possibili. Due punti fondamentali per l'esistenza di un problema decisionale sono:

- la libertà, cioè l'esistenza di più alternative disponibili;
- la razionalità, cioè l'esistenza di preferenze fra impatti distinti.

Se manca la prima, non c'è possibilità di scelta; se manca la seconda, non c'è criterio di scelta.

Può provocare qualche confusione il fatto che in informatica teorica si definisca *problema di decisione* un problema che ammette solo due possibili soluzioni (*sì e no*, oppure *vero e falso*). Si tratta di un concetto completamente diverso. Volendo trovare una relazione con i problemi dell'informatica teorica, i problemi decisionali trattati in queste note includono in qualche modo come casi particolari i *problemi di ottimizzazione/ricerca*, nei quali la soluzione è un oggetto il cui valore sia massimo (o il cui costo sia minimo) fra tutti gli oggetti di un'opportuna famiglia.

1.2 L'approccio modellistico alla decisione

La Figura 1.1 schematizza l'approccio modellistico alla decisione. Anziché passare direttamente dal problema concreto alla strategia operativa, tale approccio richiede una serie di passaggi intermedi:

- costruire un *modello* del problema, con opportuni metodi;
- risolvere il modello con *algoritmi*, cioè con metodi formali;
- interpretare la soluzione, con opportuni metodi.

In ogni fase del procedimento compaiono dei metodi. A volte si tratta di metodi formali, ovvero algoritmi. Questo riguarda specialmente la risoluzione del modello, cioè il passaggio dal modello alla soluzione astratta. La correttezza di tale passaggio può essere persino garantita da una dimostrazione matematica, se l'algoritmo è esatto. Se invece è euristico, le garanzie sono empiriche e sperimentali. Ancora più informali sono i passaggi da problema a modello, cioè dal mondo reale a quello degli oggetti matematici, e da soluzione a strategia operativa, cioè dal mondo degli oggetti matematici a quello reale. Tali passaggi sono particolarmente delicati perché intrinsecamente creativi e intuitivi, e quindi esposti a molte potenziali fonti di errore. Un rischio frequente è che i decisori confondano il modello con la realtà fidandosi ciecamente dei suoi risultati, che possono essere formalmente corretti, ma inapplicabili. Un altro rischio, più sottile, è che i decisori usino pedestremente i risultati numerici degli algoritmi per costruire una strategia operativa travisandoli (ad esempio, che applichino una soluzione “ottima”, ma molto sensibile alle variazioni dei dati, trovandosi poi a ottenere risultati pratici scadenti perché i dati reali non coincidono perfettamente con quelli usati nel calcolo). In questo caso, il modello e l'algoritmo sono corretti, ma l'interpretazione è errata.

Per questo motivo, il processo decisionale non avviene tutto in una volta, ma per correzioni successive. La Figura 1.2 illustra il procedimento schematizzato da H. A. Simon nella sua fondamentale opera “Models of bounded rationality” (1982)⁸. Lo schema presenta le seguenti fasi, nelle quali si possono riconoscere abbastanza bene i concetti introdotti più sopra:

⁸H. A. Simon (), fu uno dei fondatori dell'Intelligenza Artificiale negli anni '50 e vinse il Premio Nobel per l'Economia nel 1978 per i suoi studi sulla razionalità limitata.

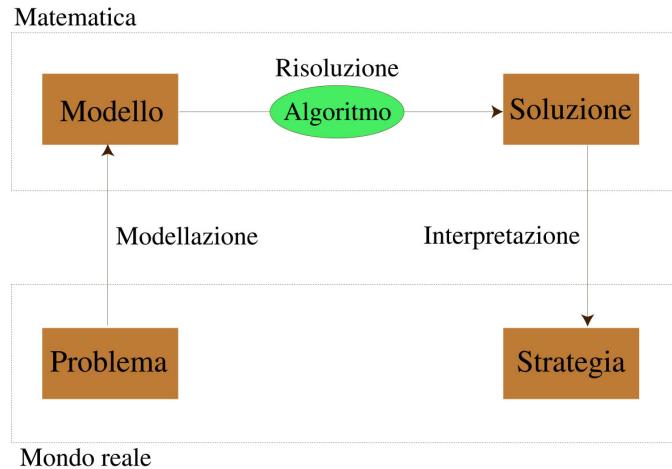


Figura 1.1: Schema dell'approccio modellistico alla decisione

1. *formulazione del problema*: si circoscrive il sistema, individuando impatti e preferenze da un lato (*Obiettivi*), decisori e scenari dall'altro (*Contesto*);
2. *identificazione delle alternative*: si definisce l'insieme delle alternative ammissibili;
3. *valutazione delle alternative e scelta*:
 - si valuta l'impatto associato ad ogni alternativa e scenario;
 - si sceglie un'alternativa in base alle preferenze dei decisori;
4. *implementazione della decisione*:
si traduce in pratica l'alternativa scelta o se ne fa una simulazione;
5. *monitoraggio e verifica*:
 - si osservano le conseguenze della decisione;
 - se insoddisfacenti, si aggiusta il processo e si ripete introducendo nuovi scenari, nuovi obiettivi, nuove alternative, nuovi metodi di valutazione.

In tutte queste fasi, sono essenziali l'esperienza specifica e l'aiuto dei mezzi informatici.

1.3 Perché un approccio formale?

È opportuno chiedersi perché sia utile formalizzare un problema decisionale. L'approccio formale è vantaggioso perché consente di:

1. prevedere in modo più sicuro e preciso l'impatto della decisione, usando modelli descrittivi anziché solo l'intuizione e l'esperienza;
2. accelerare la decisione usando algoritmi e strumenti informatici, che consentono di trattare un numero molto maggiore di alternative possibili;
3. rendicontare e certificare il processo di decisione;

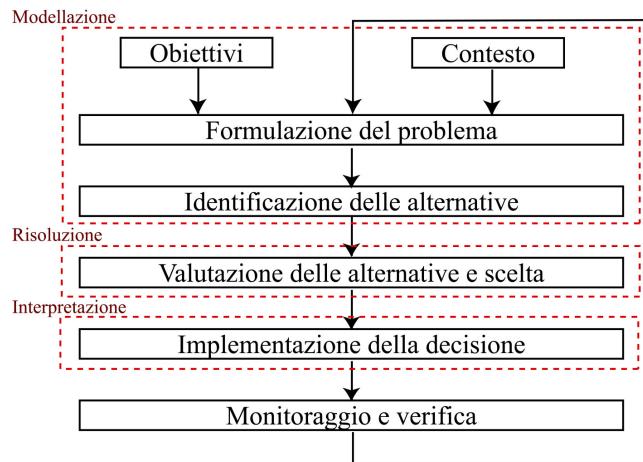


Figura 1.2: Schema del procedimento ciclico di decisione secondo Simon

- esplicitando le ipotesi fatte su alternative, scenari, preferenze e decisori e sulle loro relazioni;
- garantendo la ripetibilità del processo in modo che si ottengano sempre gli stessi risultati;
- consentendo modifiche mirate del processo senza doverlo ridiscutere in toto ogni volta che cambia qualcosa.

Si pensi soprattutto alle decisioni in ambito pubblico, oppure nelle grandi organizzazioni private: chi decide spesso non ha un potere assoluto, ma deve *dimostrare* agli altri soggetti che la propria decisione è stata presa in base ai dati disponibili e a un procedimento trasparente. In questi stessi ambiti, l'idea che, cambiata qualche circostanza secondaria, il procedimento debba ripartire da zero è forse spontanea, ma inefficiente e irragionevole.

1.4 Modelli prescrittivi e modelli descrittivi

Un modello decisionale è in genere un oggetto composto. Comprende infatti in genere un insieme di sottomodelli, che si possono classificare in due ampie categorie in base ai dati che richiedono e ai risultati che producono:

1. modelli prescrittivi (o *normativi*)

- usano come dati gli impatti per ogni configurazione possibile e le preferenze dei decisori fra gli impatti;
- danno come risultato un'alternativa “suggerita”;

In breve, *se le cose vanno così, il meglio che si può fare è questo.*

2. modelli descrittivi (o *predittivi*⁹)

- usano come dati la descrizione di ogni configurazione del sistema (scenario e alternativa);

⁹La differenza sostanziale è che i modelli predittivi sono rivolti al futuro, quelli descrittivi al presente.

- danno come risultati gli impatti che ne derivano.

In breve, *se si fa questo e succede quest'altro, allora le cose andranno così*.

I modelli descrittivi possono essere banali, ma anche sofisticati. Un esempio di modello descrittivo talmente banale da passare inosservato è quello che fornisce il costo totale c_{tot} per l'acquisto di due prodotti a partire dai loro prezzi unitari c_1 e c_2 e delle quantità acquistate x_1 e x_2 : il modello si riduce all'equazione $c_{\text{tot}} = c_1x_1 + c_2x_2$. Ci sono però modelli anche molto sofisticati, che servono a ricavare dati troppo costosi o del tutto impossibili da misurare (per esempio, previsioni del futuro, ecc...). Tutta la Parte V di queste note è dedicata alla presentazione di modelli descrittivi di un certo interesse per problemi decisionali complicati.

Le due categorie di modelli possono avere interazioni sottili. Ad esempio, qualsiasi modello prescrittivo usa un insieme di modelli descrittivi per conoscere gli impatti delle alternative e degli scenari possibili. D'altra parte, è possibile che un modello descrittivo includa al proprio interno modelli prescrittivi. Per esempio, per stimare il traffico su una strada occorre prevedere le scelte degli automobilisti di fronte a una data rete stradale, con caratteristiche fisiche, divieti e cicli semaforici dati. Tali scelte sono il risultato di decisioni, descrivibili con modelli prescrittivi. Lo stesso avviene per le scelte dei clienti al variare dei prezzi e delle condizioni di vendita di un prodotto, per quelle dei concorrenti al variare delle proprie strategie commerciali, ecc...

1.5 Business Intelligence

Business Intelligence è diventato un termine di gran moda negli studi recenti sul management di impresa. Libri e articoli di riviste la esaltano come uno strumento essenziale per vincere la competizione e avere successo. Con questo nome si intende *l'insieme dei processi e delle tecnologie usate per raccogliere, gestire e rendicontare dati orientati alle decisioni*.

L'architettura di un sistema di Business Intelligence ha lo scopo di

- portare dati, contenuti e analisi alla persona giusta nel momento giusto;
- aiutare i decisori a scegliere una soluzione, ma anche a giustificare la propria scelta, a correggerla o aggiornarla, a interagire fra loro, ecc...;
- adeguarsi alle normative sulla certificazione dei processi.

Comprende quindi sistemi tecnologici, applicazioni informatiche e procedure algoritmiche o metodologiche per eseguire le seguenti attività:

1. *Data management*: acquisire e gestire i dati;
2. *Transformation tools and processes*: estrarre, correggere, trasmettere e conservare i dati (si usa spesso l'espressione *Extract-Transform-Load*);
3. *Repository*: conservare dati e metadati;
4. *Analytics*: eseguire ottimizzazioni, simulazioni, stime, predizioni, inferenze, identificazioni di pattern e mining di informazioni;
5. *User interface*: presentare e manipolare dati e analisi senza danneggiare gli archivi dei dati;
6. *Procedure amministrative*: gestire sicurezza, errori, auditing, privacy.

Ognuno di questi punti meriterebbe un approfondimento, ma in queste note ci concentreremo sul quarto (l'*Analytics*), in particolare considerandone l'aspetto metodologico e lasciando invece in ombra quello tecnologico.

Sistemi di questo genere sono ovviamente molto complicati e contengono al loro interno sottosistemi più semplici, con compiti specifici, molti dei quali hanno una lunga storia, che risale a ben prima della moda presente. Per esempio, i sistemi di *Electronic Data Processing* (EDP) da decenni vengono usati per conservare i dati, accedervi, e risolvere semplici problemi standard. Analogamente, i *Management Information Systems* (MIS) consentono elaborazioni statistiche per aggregare dati in un formato più leggibile e immediato. Infine, i *sistemi di supporto alle decisioni* o *Decision Support Systems* (DSS) sono l'immediato predecessore dei sistemi di Business Intelligence, dato che combinano l'accesso a modelli descrittivi e prescrittivi, l'uso di sistemi logici per rivelare incoerenze, l'uso di sistemi esperti per elaborare strategie.

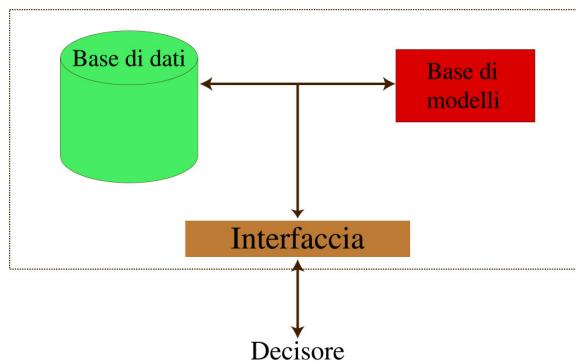


Figura 1.3: Schema generale di un sistema di supporto alle decisioni

La Figura 1.3 ne illustra uno schema generale, nel quale compaiono le principali componenti funzionali:

- una *base di dati* col relativo sistema informativo che fornisce all'utente i dati senza introdurre forzature nel loro significato;
- una *base di modelli* col relativo sistema di gestione che usa i dati per prevedere gli impatti di ogni possibile decisione e suggerirne una o più;
- un'*interfaccia utente* che consente al decisore di governare il processo, fornendo i dati in ingresso ai modelli, ricevendone i risultati ed eventualmente trasformandoli in dati per altri modelli.

In tutto questo procedimento, il gestore è responsabile di guidare ogni passo del sistema in base ai risultati intermedi, cioè di:

- definire decisori e preferenze, inserendoli come dati;
- definire alternative e scenari, inserendoli come dati;
- controllare il flusso dei dati fra i modelli, scegliendo quali dati usare e a quali modelli fornirli:
 - trasferire dati, alternative e scenari ipotetici ai modelli descrittivi, in modo da ottenere impatti simulati;

- trasferire impatti misurati o simulati ai modelli prescrittivi in modo da ottenere alternative suggerite;
- valutare se conservare i risultati dei modelli come nuovi dati;
- usare le alternative suggerite come informazione utile, non come istruzioni da eseguire.

L'aspetto fondamentale di questo procedimento è che i sistemi che lo sostengono non creano decisioni da applicare, ma solo informazioni per generare decisioni.

1.6 Esempi di problemi decisionali complicati

Questa sezione presenta alcuni esempi di problemi decisionali complicati, allo scopo di collegare i concetti precedentemente esposti con situazioni realistiche e di cominciare a suggerire quali siano gli elementi che rendono complicata in pratica una decisione. Con lo stesso scopo, il Capitolo 56 mostrerà in molto maggior dettaglio due esempi di grandi dimensioni. Dagli esempi che seguono ricaveremo un elenco dei principali aspetti complicanti, che ci porterà a classificare i problemi decisionali complicati e a costruire l'indice di queste note e del corso.

Esempio 1 (La ricerca del parcheggio) *In una città piuttosto trafficata, si sta cercando parcheggio per lasciare l'automobile e raggiungere il luogo di un appuntamento importante; si vorrebbe parcheggiare in fretta e non dover camminare molto fino alla meta.*

- *Sistema è la rete stradale locale, con l'insieme di tutti i potenziali parcheggi*
- *Alternativa è ogni possibile traiettoria (percorso e orario) per l'automobile*
- *Scenario è ogni possibile distribuzione spaziale e temporale dei parcheggi liberi*
- *Impatto sono i tempi di guida e di camminata dopo il parcheggio*
- *Decisore è il guidatore (o anche i passeggeri?)*

Si noti come l'impatto si riduca a ciò che effettivamente interessa per la decisione, ma dipenda dalla complessa combinazione (che è poi la configurazione) della traiettoria seguita dall'automobile e della disponibilità di parcheggio nei vari luoghi e momenti: la stessa traiettoria, infatti, può portare a parcheggiare immediatamente davanti al luogo dell'appuntamento o a fare lunghi giri a vuoto a seconda che il posto sia libero oppure occupato al momento in cui l'automobile ci passa.

Esempio 2 (La regolazione del termostato) *Si vuole regolare il termostato dell'aula in modo che la temperatura sia gradevole per chi tiene e per chi segue la lezione.*

- *Sistema è l'aula*
- *Alternativa è la posizione del regolatore del termostato*
- *Scenario è la temperatura esterna e l'esposizione al sole dell'aula*
- *Impatto è la temperatura interna dell'aula (ma probabilmente anche l'umidità)*
- *Decisori sono i soggiornanti nella stanza (tutti o solo il docente?)*

Per quanto questo problema sia più semplice del precedente, si noti come la definizione dell'impatto e dei decisori non sia affatto banale.

Esempio 3 (L'acquisto dell'automobile) *Si vuole comprare un'automobile che abbia buone prestazioni, sia comoda, abbia una bella linea e sia economica durante tutto il proprio ciclo di vita.*

- *Sistema è il mercato locale dell'automobile, della benzina, delle riparazioni, ecc...*
- *Alternativa è l'automobile acquistata*
- *Scenario sono le disponibilità e i prezzi dei concessionari, il verificarsi di incidenti, le tariffe della benzina e dei meccanici, ecc...*
- *Impatto sono le caratteristiche dell'automobile durante tutto il suo ciclo di vita*
- *Decisore è l'acquirente (ed eventuali familiari)*

Qui si può notare che una possibile alternativa, che è facile dimenticare, potrebbe essere di rinunciare all'acquisto dell'automobile, con un impatto ovviamente pessimo per alcuni aspetti, ma ottimo per altri.

Esempio 4 (Il turno di Risiko) *Si vuole giocare un turno di Risiko, compiendo un giro completo su tutti i giocatori.*

- *Sistema è la mappa con la distribuzione dei territori, delle armate, delle carte*
- *Alternativa sono i territori da cui e verso cui si lancia ogni attacco e il relativo numero di armate in attacco e in difesa*
- *Scenario è il risultato dei dadi ad ogni attacco*
- *Impatto è il numero di armate distrutte per ogni giocatore*
- *Decisori sono i giocatori*

Rispetto ad altri esempi in cui era possibile che ci fossero più decisori, qui la presenza di più decisori è intrinseca e ineliminabile. Inoltre, i decisori non agiscono collettivamente sulle stesse leve (nel caso dell'automobile, tutti contribuiscono a decidere l'acquisto di una sola automobile), qui ognuno compie le proprie scelte in autonomia.

1.7 Classificazione dei problemi decisionali

Gli esempi precedenti mostrano diversi fattori che possono complicare un problema decisionale. Cominciamo col dire che non ci occupiamo qui della così detta *complessità computazionale*, cioè del fatto che esistono problemi perfettamente chiari e definiti per i quali non è noto alcun algoritmo polinomiale e si congettura che esistano solo algoritmi esponenziali¹⁰. In questi problemi è costoso passare dal modello alla soluzione, ma non è complicato definire quale sia la soluzione corretta.

I fattori che ci interessano riguardano invece i decisori e le loro relazioni col sistema. Più precisamente:

¹⁰Per distinguersi da questo concetto, quasi sempre useremo le espressioni “complicato” e “complicazione” anziché “complesso” e “complessità”. Sfortunatamente, fra poche righe parleremo di “preferenze complesse”, perché “preferenze complicate” suona proprio male.

1. *preferenze complesse*, cioè insufficienti a definire un concetto di ottimo;
2. *scenari molteplici*, dunque un *ambiente incerto*;
3. *decisori molteplici*, con preferenze potenzialmente in conflitto tra loro.

C'è un ulteriore fattore che può complicare un problema di decisione, ed è una *formalizzazione insufficiente* del sistema, ovvero l'incapacità di determinare in modo univoco l'impatto per una data configurazione, cioè avendo fissato alternativa e scenario. Questa situazione viene però spesso ricompresa nel caso di ambiente incerto, ipotizzando che, oltre alle componenti note dello scenario, vi siano altre componenti aggiuntive incognite (a volte definite *disturbi*), note le quali sarebbe possibile calcolare esattamente l'impatto. I modelli che ne derivano sono detti *modelli a scatola nera*, o *modelli a scatola grigia*, a seconda di quanto sono importanti i disturbi rispetto alle componenti misurabili della configurazione.

A questo punto, si possono classificare i problemi decisionali in base al fatto che ciascuno dei tre principali fattori complicanti sia presente oppure no, ottenendo $2^3 = 8$ famiglie di possibili problemi decisionali. La Figura 1.4 descrive questa classificazione con uno schema a tre assi.

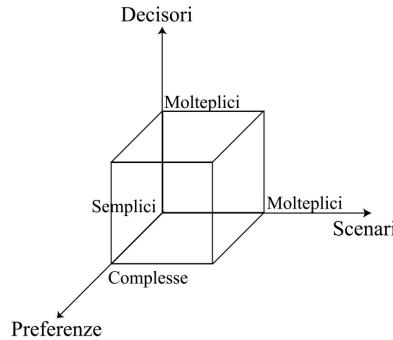


Figura 1.4: Classificazione dei problemi decisionali in base ai tre fattori complicanti

Sullo schema a tre assi si basa il programma del corso e l'indice di queste note. Invece di trattare tutte le 8 classi, ci limiteremo a considerare le 4 famiglie fondamentali in cui nessuno o uno solo dei 3 fattori complicanti si presenta, discutendo eventuali sottofamiglie importanti. Le altre 4 famiglie si possono trattare con opportune combinazioni dei concetti adottati per le famiglie di base. Il corso e queste note quindi presentano nelle prime quattro parti altrettante famiglie di modelli prescrittivi:

1. modelli con preferenza, scenario e decisore semplici (*Programmazione Matematica*)
2. modelli con preferenze complesse, un solo scenario, un solo decisore:
 - obiettivi molteplici (*Programmazione a molti obiettivi*)
 - preferenza non modellata da obiettivi (*Analisi gerarchica*)
 - preferenza non transitiva (*Metodi Electre*)
3. modelli con preferenze semplici, scenari molteplici, decisore semplice:
 - incertezza assoluta (*Decisioni in condizioni di ignoranza o Programmazione robusta*)

- incertezza descritta probabilisticamente (*Decisioni in condizioni di rischio* o *Programmazione stocastica*)
4. modelli con preferenza semplice, scenario semplice, decisori molteplici:
- decisori indipendenti (*Teoria dei giochi*)
 - decisori costretti a coordinarsi (*Decisioni di gruppo*)

e nella quinta parte alcune famiglie esemplari di modelli descrittivi sofisticati:

- teoria delle code
- simulazione a eventi discreti
- modelli di trasporto e di scelta discreta
- modelli dinamici

Capitolo 2

Casi di studio

Questo capitolo è dedicato a presentare due casi di studio che chiariscano il concetto di problema decisionale complicato e facciano toccar con mano il significato concreto delle parole chiave presentate nel Capitolo 1. Lo scopo dei casi di studio non è di essere studiati nel dettaglio per passare l'esame, ma di essere usati per costruirsi un'idea concreta delle definizioni astratte alla base del corso e per imparare a riconoscere tali concetti in altre situazioni, anche molto diverse. Va anche precisato che si tratta di studi relativi a opere non realizzate, e quindi potenzialmente soggetti a innumerevoli varianti e correzioni. Il primo caso, in particolare, è uno studio assolutamente preliminare condotto nell'ambito di una tesi di laurea con la collaborazione del Comune di Como, mentre il secondo è uno studio di fattibilità ufficiale affidato dal Comune di Milano al Dipartimento di architettura e studi urbani del Politecnico di Milano.

Entrambi i casi di studio si riferiscono a grandi opere pubbliche. Questo è solo uno dei campi in cui si possono presentare decisioni complicate, ma si presta bene a fornire esempi didattici sofisticati perché:

1. informazioni dettagliate sono più spesso disponibili pubblicamente;
2. i tre fattori complicanti elencati nell'introduzione (preferenze complesse, incertezza e attori molteplici) sono presenti al massimo grado.

Riguardo il primo punto, spesso la legge prescrive che informazioni sufficienti siano diffuse pubblicamente affinché i cittadini possano farsene un'idea completa e corretta. D'altra parte, spesso le istituzioni decidono di farlo spontaneamente per informare l'opinione pubblica o per acquistarne il consenso. Nel caso di decisioni aziendali o comunque private, questo è molto più raro. Per quanto riguarda il secondo punto, le decisioni sulle grandi opere pubbliche hanno per propria natura impatti molteplici (economici, sociali, ambientali, ecc...), e quindi è difficile esprimere la preferenza come la semplice minimizzazione di un costo o massimizzazione di un profitto. D'altra parte, tali decisioni dipendono anche da eventi non controllabili dai decisori (fenomeni sociali o economici su scala più ampia, eventi politici, ecc...). Infine, esse coinvolgono grandi quantità di persone, associazioni e istituzioni diverse che, se pure non sono incaricate della decisione, possono esercitare un ruolo in essa o una pressione sui decisorи.

Tutto questo giustifica la scelta di esempi tratti da tale campo.

2.1 La Metrotranvia di Como

2.1.1 Il contesto

Como è da secoli al centro di traffici commerciali e spostamenti di persone. Si trova infatti all'incrocio fra una direttrice principale (nord-sud), che lega l'Italia (in particolare Milano) alla Svizzera e alla Germania, e una direttrice secondaria (est-ovest), che lega i centri principali all'imbocco delle valli alpine (Varese, Lecco, Bergamo, ecc...). La Figura 2.1 mostra queste due direttrici.

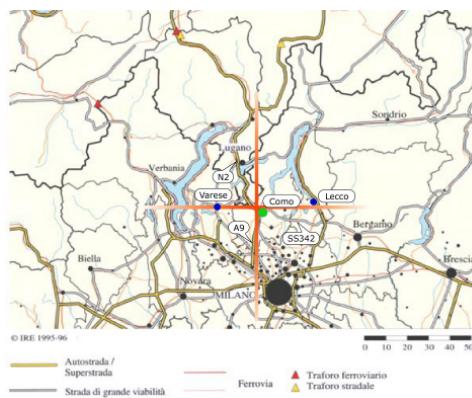


Figura 2.1: Le due principali direttrici di traffico che si incontrano in Como

La posizione geografica che ha favorito lo sviluppo della città è all'origine di alcuni dei problemi che ora la affliggono, e che sono legati essenzialmente alla saturazione delle vie di accesso, con le sue naturali conseguenze:

- traffico (vedi le Figure 2.2 e 2.3),
- perdite economiche dovute alla lentezza degli spostamenti,
- inquinamento (vedi la Figura 2.4),
- incidenti.

I mezzi privati sono i principali responsabili di tali problemi, dato che si concentrano su poche direttrici di ampiezza piuttosto scarsa (vedi la Figura 2.5). La principale è l'accesso da sud-ovest attraverso la via Napoleona che raccoglie le stazioni Briantea e dei Giovi e la provinciale Canturina. Seguono le statali Regina e Lariana, a nord, strette sulle due sponde del lago, e la statale Briantea a est.

Allargare questi accessi o crearne altri è complicato e costoso per la struttura orografica del territorio: Como giace in una stretta piana lungo il lago, circondata da alte colline. Solo la stretta breccia di Grandate la collega alla pianura a sud (vedi la Figura 2.6).

Un possibile miglioramento sarebbe trasferire spostamenti sul mezzo pubblico, che è più efficiente di quello privato dal punto di vista della capacità di trasporto. Tuttavia, le indagini mostrano che la percentuale di spostamenti serviti dal mezzo pubblico è nettamente inferiore, e soprattutto è diseguale. Le autolinee servono una minoranza di utenti, tendenzialmente in calo. D'altra parte, sono soggette alle stesse strozzature dei mezzi privati. Per quanto riguarda le ferrovie (vedi Figura 2.7), mentre la linea statale attestata a Como S. Giovanni (ex-FS) è relativamente scarsa, quella delle ferrovie regionali attestata a Como Lago (ex-FNM) è vicina a

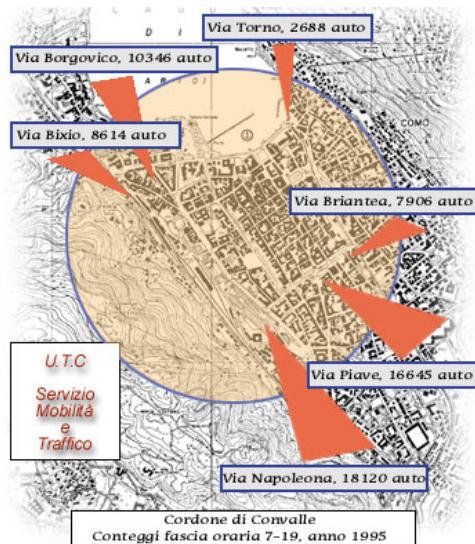


Figura 2.2: Flussi veicolari quotidiani in ingresso alla città di Como

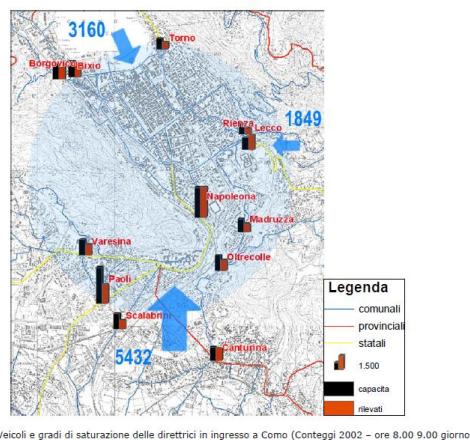
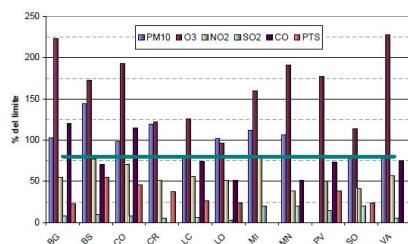


Figura 2.3: Livelli di saturazione delle vie d'accesso a Como nelle ore di punta mattutine



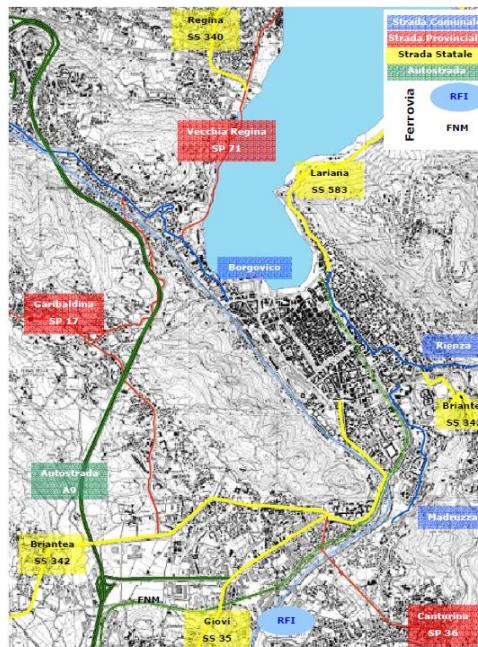


Figura 1.1: Le direttrici automobilistiche di accesso.

Figura 2.5: Le principali vie d'accesso a Como per il traffico automobilistico

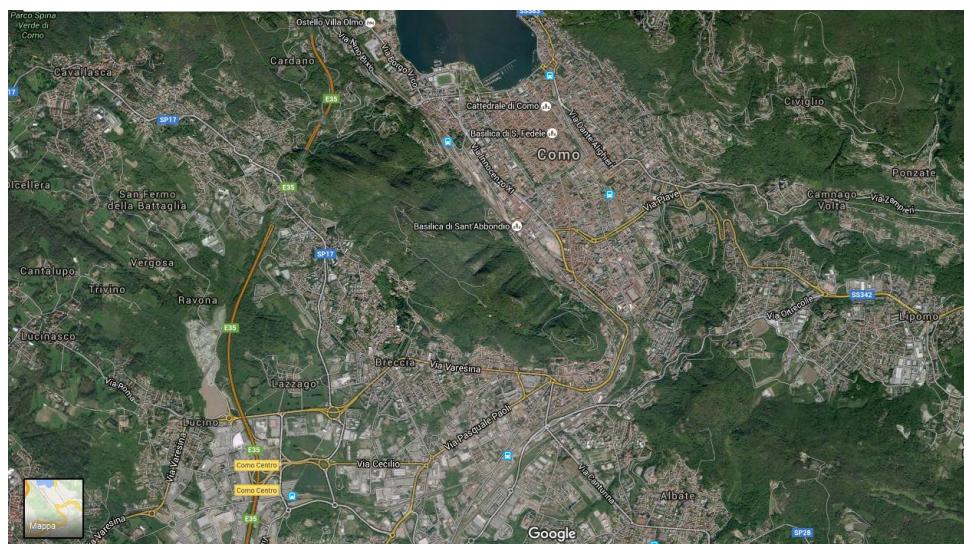
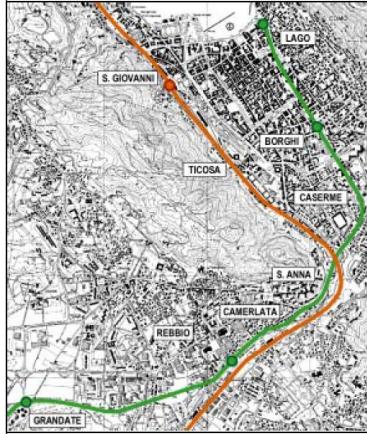


Figura 2.6: Orografia dell'area di Como

saturazione¹. Viene quindi spontaneo pensare di potenziare la capacità della linea FNM ed eventualmente spostare sulla linea FS parte della relativa domanda. Come fare?



Servizio ferroviario attuale in Como (in verde la linea FNM e in arancione quella FS).

Figura 2.7: Le due linee di accesso ferroviario a Como: in arancione quella statale e in verde quella regionale

Per discutere meglio le interazioni potenziali fra le due linee, occorre elencare alcuni punti critici:

1. l'incrocio fra le linee FS e FNM ha luogo in piena breccia, con un acrobatico sistema di ponti: la via Napoleona passa sopra la linea FNM che passa sopra la linea FS, che passa sopra una piccola strada locale, via Mulini (vedi le Figure 2.8 and 2.9);



Figura 2.8: La linea FS passa sopra via Mulini

¹Dal momento dello studio, le ferrovie regionali hanno cominciato a fare servizio anche sulla linea statale, con treni da Milano a Como San Giovanni. Questo ha leggermente migliorato la situazione, sfruttando meglio la capacità residua della linea più scarica. Inoltre, le due compagnie hanno cambiato nome. Per semplicità, però, continueremo a indicarle con i loro nomi storici.



Figura 2.9: La linea FNM passa sopra la linea FS

2. fra le stazioni di Camerlata e Borghi, la linea FNM è a binario singolo, dato che occupa uno stretto terrapieno con ripide scarpate su entrambi i lati;
3. fra la stazione di Borghi e quella di Lago, la linea FNM è a binario singolo e passa in mezzo a uno stretto budello fra le case (vedi Figura 2.10);
4. se si volessero collegare le due linee attraverso la piana della città, bisognerebbe passare o entro le mura (in zona storica) o lungo il lago (in zona turistica, occupata da una strada principale e di tanto in tanto soggetta a esondazione) e poi risalire il ripido pendio in cima al quale sorge la stazione FS di Como S.Giovanni.

Tutti questi fattori rendono una soluzione ferroviaria irrealizzabile, a meno di limitarsi a usare solo le tratte esistenti. Incrementare la frequenza dei treni sulla linea FNM però presenta due problemi:

- la linea a binario singolo impedisce a due treni di percorrere insieme la tratta Camerlata-Borghi-Lago (a meno di realizzare incroci perfettamente sincronizzati a Como Borghi, tecnicamente possibili, ma destinati a creare problemi ad ogni minimo imprevisto);
- i passaggi a livello tagliano la città in due per diversi minuti ad ogni passaggio di treno: se passa un treno ogni 10-15 minuti anziché ogni mezz'ora, l'impatto sul traffico diventa molto pesante.

Di qui l'idea di una tranvia:

- realizzabile a doppio senso, dato lo scartamento ridotto;
- con incroci semaforici veloci anziché passaggi a livello;
- prolungabile entro la città.

Ma una tranvia classica ha il problema di richiedere la sostituzione dell'intera linea, dunque:

- costringere i passeggeri in arrivo da fuori a cambiare in una stazione esterna;
- interrompere il traffico nel corso dei lavori.



Figura 2.10: La linea FNM fra le stazioni di Como Borghi e Como Lago

2.1.2 La generazione delle alternative

Contrariamente ai modelli base della Ricerca Operativa, o ai problemi anche complessi che sorgono in ambiti ben definiti e ristretti, qui non è affatto chiaro quale sia l'insieme delle possibili alternative: verrebbe da dire che sia infinito. È più corretto dire che è *indefinito*, e che va costruito un po' per volta studiando il contesto. Inoltre, può essere costruito per iterazioni successive. Come si diceva nell'introduzione:

- si ipotizzano alcune alternative
- se ne valutano gli impatti
- se nessuna è soddisfacente, si generano nuove alternative, guidati dalle informazioni ricavate nello studio, e si torna a capo.

Per esempio, se ci si rende conto che cambiare i binari costa troppo, ci si può chiedere se non esistano tram capaci di usare i binari del treno. Guarda caso, esistono².

Il processo di generazione delle alternative è facilitato se si ragiona scomponendo il problema, cioè individuando gli *elementi di alternativa* che riguardano ambiti diversi e (almeno in parte) indipendenti. In questo studio si sono individuati tre elementi, sui quali fondare la generazione delle alternative:

1. la tecnologia usata per realizzare la nuova linea;
2. il percorso della nuova linea;
3. gestione dei treni FNM rispetto alla nuova linea.

Discussiamo in ulteriore dettaglio ciascuno di questi elementi.

²Si veda <https://it.wikipedia.org/wiki/Tram-treno>

Tecnologia Studi precedenti suggeriscono tre tipologie tecnologiche:

1. servizio ferroviario: si aggiunge un treno-navetta fra Grandate e Como Lago, così da aumentare la frequenza delle corse dell'attuale linea ferroviaria;
2. metrotranvia: si sostituisce la ferrovia con una metrotranvia classica;
3. interoperabile: si fa coesistere il servizio ferroviario con una metrotranvia che usa mezzi capaci di muoversi nel doppio ambiente (con i relativi problemi di alimentazione, di scartamento dei binari, ecc...).

Tracciato Il numero di alternative è sterminato. Per cominciare, si sono individuate quattro possibilità (vedi la Figura 2.11):

1. mantenimento dell'attuale tracciato FNM, con termine del servizio alla stazione di Como lago;
2. collegamento delle stazioni Como Lago (FNM) e Como S. Giovanni (FS) attraverso binari passanti per l'interno della città;
3. collegamento delle stazioni Como Lago (FNM) e Como S. Giovanni (FS) attraverso un tratto di binari passanti per il lungolago;
4. chiusura ad anello del centro cittadino attraverso la realizzazione del tratto Como Lago - Como S. Giovanni, più un collegamento diretto Como Borghi (FNM) - Como S. Giovanni.

Gestione dei treni FNM Con questo elemento di alternativa si intende la scelta di come far convivere i due sistemi di trasporto su ferro:

1. mantenere le FNM sul loro tragitto attuale;
2. deviare il traffico FNM sulla stazione FS di Como S. Giovanni, attraverso la costruzione di un interscambio fra le linee FNM e FS (probabilmente nella stazione di Camerlata).

Si tratta di due ipotesi estreme, fra le quali esistono molte possibilità intermedie, sia durante i lavori (dirottare temporaneamente le corse FNM su Como S. Giovanni nei periodi in cui la tratta è bloccata per lavori) sia in permanenza. Per esempio, oggi le FNM gestiscono anche il traffico locale diretto a Como S. Giovanni, anche se le due linee rimangono completamente separate.

Ogni aspetto costituisce un *elemento di alternativa*. Le singole scelte possibili per un elemento possono essere associate a valori numerici: a volte si tratterà di una vera e propria misura quantitativa, altre volte di un semplice indice arbitrario. Ad esempio, si può usare 0 per il treno, 1 per il tram, 2 per l'interoperabile. Lo scopo di questa associazione, in generale, non è di fare conti, ma solo di descrivere in modo efficace l'insieme delle alternative.

Una volta individuati gli elementi di alternativa e, per ciascuno, i possibili valori, si può procedere a elencarne le combinazioni, che in generale non sono tutte possibili. Nel caso in esame, le possibili combinazioni sono $3 \times 4 \times 2 = 24$, ma solo 6 sono realizzabili. Le altre possono essere escluse a causa di:

- ovvie contraddizioni;

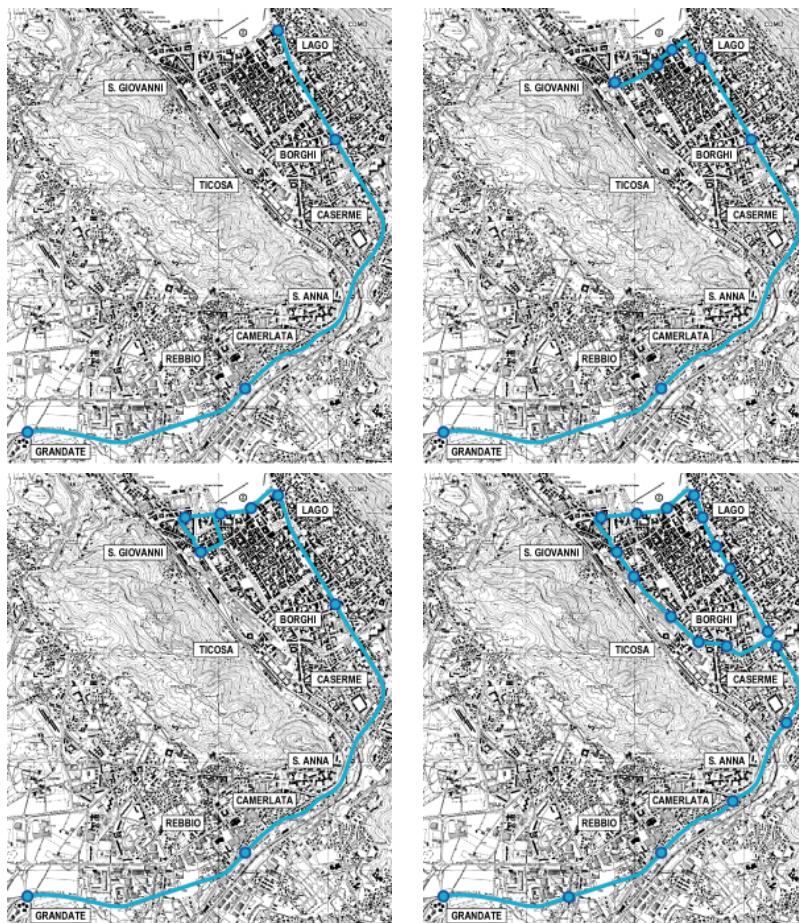


Figura 2.11: I quattro percorsi alternativi per il progetto della Metrotranvia di Como

- impossibilità tecnica o costi troppo elevati;
- buon senso.

Per esempio, la tecnologia ferroviaria è praticamente incompatibile con qualsiasi percorso diverso dall'attuale: oggi non si possono più far passare i treni in una città come si faceva fino ai primi del Novecento, salvo in galleria, ipotesi qui improponibile. La Figura 2.12 mostra le possibili combinazioni e le alternative ammissibili per il caso di studio.

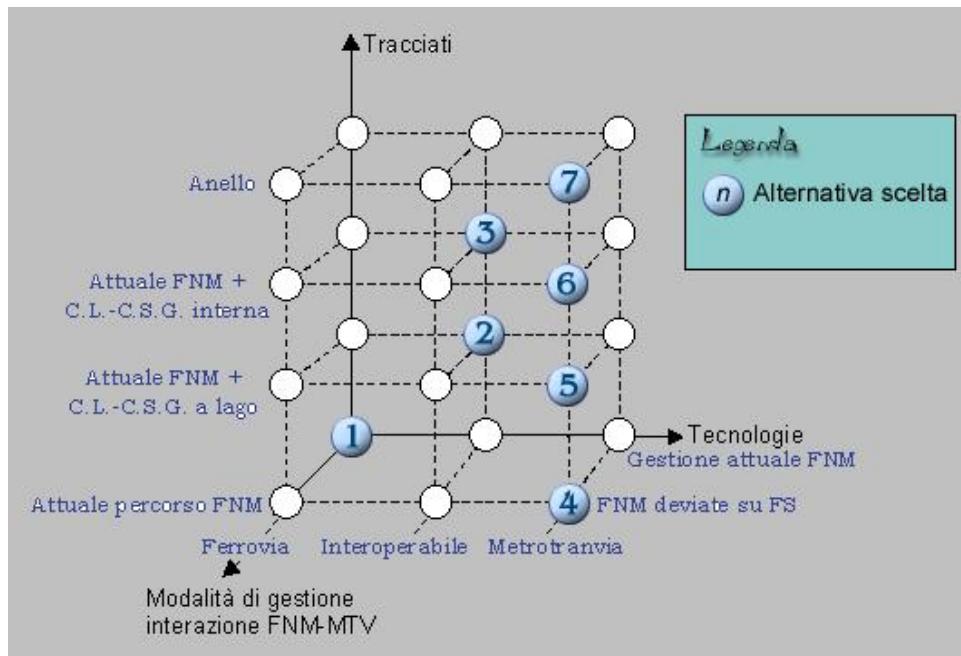


Figura 2.12: Le 24 combinazioni e le 7 alternative ammissibili per il caso di studio sulla Metrotranvia di Como; i tre assi rappresentano gli elementi di alternativa

Un punto fondamentale nel meccanismo di generazione delle alternative è che *esiste sempre un'alternativa che consiste nel non realizzare alcun intervento e conservare la situazione attuale*. Questa alternativa si indica convenzionalmente con il termine *alternativa zero*³. Questa osservazione è importante, perché spesso la foga del dibattito pubblico fa dimenticare che eventuali soluzioni alternative devono, come minimo, non peggiorare la situazione. E siccome “peggiорare” è un concetto complesso, questo punto è tutt’altro che banale.

In questa trattazione abbiamo ignorato altri potenziali elementi di alternativa, come:

- la realizzazione di parcheggi di interscambio (in vari punti);
- la realizzazione di zone a doppio binario lungo la tratta a binario unico;
- la realizzazione di nuove fermate lungo il tracciato attuale e le eventuali tratte aggiuntive;
- eventuali spostamenti del tracciato attuale;

³Nello caso di studio, purtroppo, è indicata con l'indice 1.

- il prolungamento del servizio fuori città oltre Grandate;
- la frequenza della nuova linea;
- le tariffe della nuova linea;
- ...

In particolare, va citato il fatto che studi sulla mobilità suggeriscono che ci possa essere una forte richiesta per una linea che serva zone della provincia attualmente mal collegate a Como, sia verso est (Cantù) sia ovest (Olgiate Comasco). Studi successivi a quello usato per questa dispensa, infatti, si sono concentrati su questi aspetti nella generazione delle alternative, mentre hanno ridotto i percorsi cittadini scartando le ipotesi estreme (percorso attuale e anello cittadino) e fondendo quelle intermedie (attraversamento del centro e passaggio sul lungolago) in una variante mista che passa prima per il lungolago e poi attraversa una parte del centro (vedi la Figura 2.13).



Variante Mista, il collegamento Como San Giovanni (FS) - Como Lago (FNM).

Figura 2.13: L’alternativa di percorso mista che passa prima per il lungolago e poi attraverso il centro

Abbiamo inoltre ignorato le possibili *misure di mitigazione*, cioè quegli interventi che vanno aggiunti alle alternative per ovviare agli impatti negativi che esse producono (deviazioni del traffico o delle corse ferroviarie, opere di arredo urbano per ridurre l’impatto della nuova linea sul percorso, ecc...) o per convincere eventuali attori fortemente contrari a condividere o almeno non osteggiare fortemente il progetto (ad es., i parchi, le scuole, gli impianti sportivi che vengono donate ai comuni nel cui territorio si costruisce un’autostrada, una discarica, una centrale energetica, ecc...) o infine per modificare gli impatti dell’opera (ad es., zone a

traffico limitato per ostacolare il mezzo privato e parcheggi di corrispondenza per invogliare i viaggiatori a usare il nuovo mezzo pubblico più di quanto non farebbero spontaneamente).

Tutto ciò conferma che già solo elencare le soluzioni possibili di un problema decisionale può essere un problema preliminare, che non si risolve in un solo passo, e in effetti non è mai risolto in maniera definitiva.

2.1.3 La generazione degli scenari

La stessa complicazione riguarda la definizione degli scenari, con in più la necessità di prevedere eventi incontrollabili dai decisorì, ma con un impatto potenziale sull'opera.

La generazione degli scenari segue la stessa logica della generazione delle alternative: si individuano gli *elementi di scenario* con i loro possibili valori, si elencano le loro combinazioni e si escludono quelle considerate impossibili. Ovviamente, questo esclude dallo studio ogni elemento non previsto e ogni combinazione esclusa, una cosa che mina alla base il risultato. Del resto, l'alternativa è procedere a caso o per intuito, che sarebbe peggio. Come si è già detto, se nuovi scenari emergono o se ne svaniscono di vecchi, lo studio va semplicemente aggiornato e adattato, eventualmente col risultato di suggerire un'alternativa differente.

Nel caso in esame, gli elementi di scenario considerati sono i seguenti:

1. chiusura al traffico privato del lungolago: si tratta di chiudere l'accesso ai veicoli della strada che passa fra il centro di Como e il lago: questo interagisce con il progetto rendendo più facile da realizzare una tranvia che passi lungo il lago (altrimenti, lo spazio sarebbe ridotto);
2. Stazione Internazionale Como-Chiasso: si tratta dell'unificazione delle stazioni internazionali di Como e di Chiasso in una sola fermata, a metà strada (vedi la Figura 2.14), accelerando gli spostamenti fra Italia e Europa centrale: questo interagisce con il progetto rendendo utile una linea che porti utenti alla stazione di Como S. Giovanni (declassata a locale) e magari proseguia fino alla nuova stazione internazionale⁴
3. Paratie antiesondazione: si tratta di costruire barriere (in parte mobili e in parte fisse) che proteggano il centro di Como dalle esondazioni del lago (vedi Figura 2.15): questo interagisce col progetto favorendo il passaggio sul lungolago di una nuova linea che altrimenti sarebbe resa inoperante dalle esondazioni⁵
4. Tunnel del Borgovico: si tratta di un tunnel a pedaggio che dovrebbe correre dalla zona nord-ovest alla zona sud-ovest di Como consentendo l'attraversamento in direzione nord-sud senza intasare la parte occidentale della città⁶
5. Tunnel sotto il lago: si tratta di un tunnel che dovrebbe correre sotto il lago (vedi la Figura 2.16) sostituendo il lungolago nel collegamento della zona nord-ovest con la zona nord-est di Como⁷

⁴La stazione è ancora un progetto; la tendenza attuale è a sopprimere Como a favore di Chiasso.

⁵Le paratie sono in lenta costruzione, con forti polemiche sul loro impatto visivo. Un tratto di paratia fissa è stato persino costruito e poi abbattuto perché chiudeva la vista sul lago ai passanti...

⁶Progetto apparentemente accantonato in quanto faraonico.

⁷Progetto apparentemente accantonato in quanto faraonico.

Teoricamente, ogni singola combinazione di questi elementi è possibile, salvo quelle che combinano il tunnel sotto il lago con la non chiusura del lungolago (ipotesi insensata) e quelle che combinano entrambi i tunnel (costi eccessivi).



Figura 2.14: La Stazione Internazionale Como-Chiasso

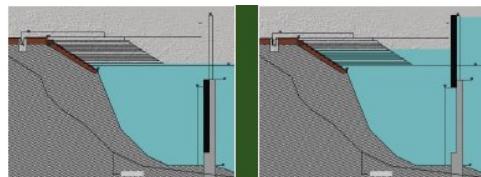


Figura 2.15: Le paratie antiesondazione lungo il lago

In tutto questo mancano aspetti importanti, come

- le variazioni presunte nella struttura residenziale e produttiva dell'area di studio;
- le variazioni della *matrice origine destinazione* (*matrice OD*) della domanda potenziale (vale a dire il numero di spostamenti da luogo a luogo che avverranno e che potrebbero essere catturati dalla nuova linea);
- l'entità di finanziamenti europei, statali, regionali all'opera.

Alcuni di questi elementi di scenario corrispondono a valori numerici, che possono essere compresi in intervalli prevedibili. Altri invece possono corrispondere a

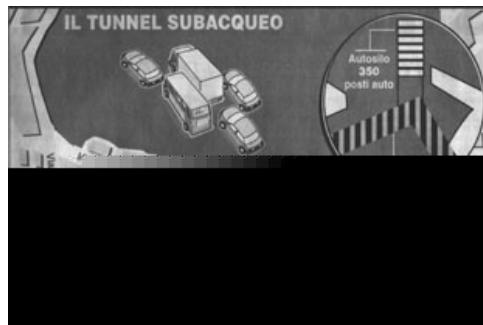


Figura 2.16: Il tunnel sotto il lago

eventi che potrebbero verificarsi o no (o verificarsi in modi diversi). Per ciascuno, può essere stimabile una probabilità, o quanto meno si può dare una valutazione qualitativa della verosimiglianza che si realizzi. Tutto questo va a descrivere l'insieme degli scenari, con le eventuali informazioni accessorie.

2.1.4 La definizione e il calcolo degli indicatori

Con *indicatori* si intendono gli *elementi di impatto*, cioè tutte quelle grandezze che caratterizzano il gradimento dei decisorì per una configurazione del sistema, e quindi influenzano le loro preferenze.

Gli indicatori tipicamente sono ancora più numerosi e diversificati degli elementi di alternativa e di scenario. Lo sono a tal punto che, per evitare di dimenticare degli indicatori rilevanti, la generazione degli indicatori viene svolta dettagliando progressivamente i diversi ambiti dell'impatto con un procedimento gerarchico:

- si individuano dei macrosettori molto generali;
- si scompono ogni macrosettore in settori, ed eventualmente questi in sottosettori, via via più specifici;
- infine, si individuano i singoli indicatori elementari.

Questo meccanismo produce un vero e proprio *albero degli indicatori*.

Nel caso in esame, si sono adottati i tre macrosettori standard in uso nelle *Valutazioni di Impatto Ambientale* (VIA)

1. *Ambiente*, a sua volta suddiviso in Aria, Rumore, Vibrazioni, Paesaggio e Assetto territoriale (cioè la compatibilità con i piani di gestione del territorio in vigore)⁸;
2. Economia, suddivisa in Costi, Ricavi e Valore immobiliare
3. Società, suddivisa in Apprezzamento, Disagi, Accessibilità⁹, Occupazione, Effetti indotti

a cui si è aggiunto il macrosettore

⁸L'inquinamento di Acqua e Suolo sono stati ignorati, dato che l'impatto della Metrotranvia sarà probabilmente molto limitato da questi punti di vista.

⁹Questo indicatore misura il tempo e il costo richiesti per accedere a luoghi importanti della città, come ospedali, monumenti, scuole, ecc... .

4. Trasporti, suddiviso in Sicurezza, Congestione e Interferenze

Quest'ultimo macrosettore in genere ricade sotto quello della Società, ma in questo caso gli si è attribuito un ruolo autonomo, trattandosi della valutazione di una grande opera trasportistica. Tutto ciò ha dato luogo all'albero degli indicatori riportato nella Figura 2.17.

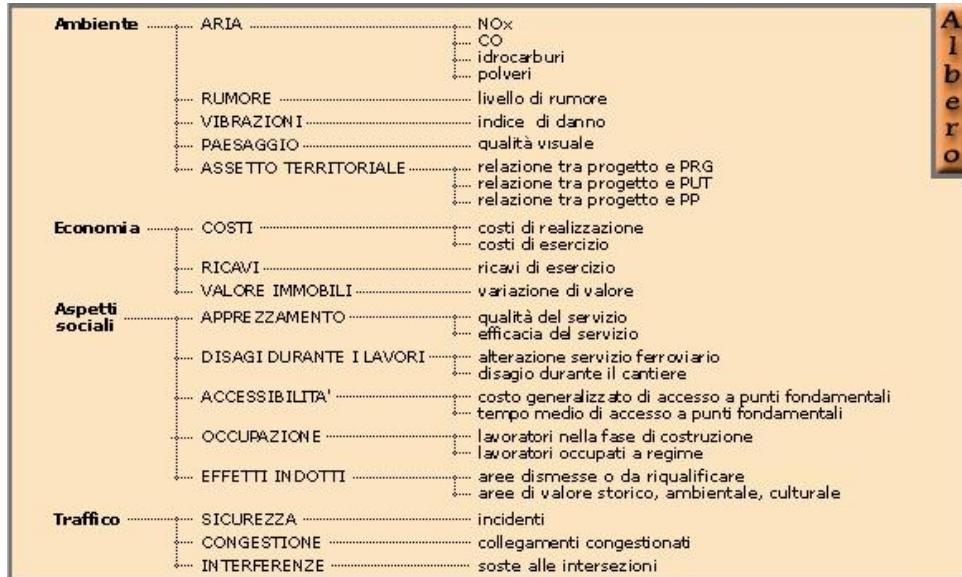


Figura 2.17: L'albero degli indicatori

Lo scopo della struttura gerarchica è anche di organizzare la valutazione dei singoli indicatori nella fase di raccolta dei dati. La stima e la previsione del valore di indicatori di un settore verranno infatti assegnate a esperti del settore stesso: esperti ambientali si occuperanno di valutare i livelli di inquinamento previsti per ogni alternativa e scenario, esperti di contabilità i costi di realizzazione e gestione, esperti di urbanistica la qualità visiva dell'opera, esperti di trasporti l'impatto sul traffico urbano e sui tempi di spostamento, ecc...

Allo stesso modo, la costruzione delle preferenze nella fase di decisione vera e propria sarà organizzata in base alle competenze specifiche degli esperti dei vari settori. La scelta se considerare l'inquinamento da ossidi di carbonio più o meno importante di quella da ossidi di azoto sarà devoluta a esperti di medicina, quella del peso relativo fra costi di realizzazione e di esercizio (l'ammortamento, diciamo) è assegnata agli esperti di economia. La struttura ad albero implica che ai livelli superiori, i decisori si occupino di definire il peso relativo fra inquinamento dell'aria e rumore, sicurezza stradale e congestione stradale, e infine fra i macrosettore Ambiente, Economia, Società e Trasporti. Questi concetti sempre più astratti comportano competenze via via sempre più "politiche". Quindi nessuno sarà mai chiamato a confrontare indicatori del tutto eterogenei, come l'inquinamento da CO_2 con il risparmio di tempo nel raggiungere la stazione di Como Lago, o la qualità visuale della tranvia sul lungolago con il disagio prodotto dai cantieri durante la sua costruzione. Si confronteranno invece CO_2 con NO_x , costi di viaggio con tempi di viaggio, l'aderenza al Piano Urbano del Traffico con quella al Piano Regolatore Generale. Rimandiamo a dopo il problema di come sia possibile confrontare fra loro termini astratti e vaghi come i macrosettori.

Ogni indicatore va associato con lo strumento necessario per ottenerne il valore in corrispondenza a ciascuna alternativa e a ciascuno scenario. Siccome nella maggior parte dei casi si parla di valori non misurabili, perché riferiti a una situazione ipotetica (l'inquinamento da CO_2 dell'alternativa zero nello scenario attuale è misurabile direttamente, ma quello dell'alternativa tranvia ad anello nello scenario in cui si chiude il lungolago al traffico, si fanno le paratie antiesondazione e basta va invece stimato in qualche modo. Molto spesso, quindi, il valore dell'indicatore è il risultato di un modello descrittivo, del quale occorre individuare i dati necessari.

Gli indicatori possono essere valori numerici che esprimono misure fisiche, ma anche valori qualitativi espressi su una scala ordinale (ad es., la qualità visuale può essere scadente, mediocre, sufficiente, discreta, buona, ottima). Questi valori si possono tradurre in una scala numerica per brevità. Anche qui, lo scopo (almeno all'inizio) non è di fare dei calcoli, ma di descrivere la situazione. In realtà vedremo che alcuni metodi di valutazione effettivamente si serviranno di tali valori numerici per fare veri e propri calcoli.

Decomposizione temporale (fasi) e spaziale (zone) Di alcuni indicatori è necessario precisare l'aspetto *spaziale* e *temporale*. In altre parole, può avvenire che non abbia senso definire il valore di un indicatore globalmente sull'intero territorio o sull'intero orizzonte temporale dell'opera.

In tal caso, per descrivere correttamente la situazione occorre suddividere il sistema in zone geografiche distinte, su ognuna delle quali abbia senso, almeno approssimativamente, definire un valore per l'indicatore considerato. Per esempio, il tempo richiesto a raggiungere la stazione di Como Lago dipenderà fortemente dal punto di partenza, anche se si può considerare approssimativamente uniforme per una data zona della città. È chiaro che questo aumenta il numero delle informazioni da gestire, e richiederà di valutare anche un'importanza relativa delle diverse zone sulla decisione, trattando i valori degli indicatori nelle zone come se fossero valori di indicatori diversi. Altri indicatori, pur georeferenziati, si possono mantenere aggregati, in quanto il loro effetto riguarda l'intera città (ad es., il danneggiamento degli edifici monumentali, che sono patrimonio comune). Si veda nella Figura 2.18 la distinzione tra indicatori aggregati e disaggregati per zone nel caso di studio.

Nel caso di studio, si è adottata una suddivisione in 5 zone già utilizzata per precedenti lavori sul territorio del Comune di Como (vedi la Figura 2.19):

1. Como centro
2. Borghi
3. Camerlata
4. Lora
5. Tavernola

Solo le prime tre zone sono direttamente influenzate dal progetto in esame, mentre Lora e Tavernola sono state mantenute per verificare gli effetti indiretti su tutto l'ambito comunale.

2.1.5 La definizione degli attori

Con *attore* si intende qualsiasi persona, organizzazione, categoria o istituzione sia coinvolta, in modo diretto o indiretto, dal progetto, anche se non ha un ruolo nella decisione, dato che, essendo i suoi interessi influenzati dal progetto, è prevedibile

SETTORI	SOTTOSETTORI	INDICATORI	DISAGGREGATI (PER ZONE)	AGGREGATI
Ambiente	ARIA	NO _x	X	
		CO	X	
		idrocarburi	X	
		polveri	X	
	RUMORE	livello di rumore	X	
	VIBRAZIONI	indice di danno		X
	PAESAGGIO	qualità visuale		X
	ASSETTO TERRITORIALE	relazione tra progetto e piani		X
	COSTI	costi di realizzazione		X
		costi di esercizio		X
Economia	RICAVI	ricavi di esercizio		X
	VALORE IMMOBILI	variazione di valore		X
	APPREZZAMENTO	qualità del servizio		X
		efficacia del servizio		X
	DISAGI DURANTE I LAVORI	alterazione servizio ferroviario		X
		disagio durante il cantiere	X	
	ACCESSIBILITA'	costo generalizzato di accesso ad alcuni punti fondamentali		X
		tempo medio di accesso ad alcuni punti fondamentali		X
	OCCUPAZIONE	lavoratori nella fase di costruzione		X
		lavoratori occupati a regime		X
Aspetti sociali	EFFETTI INDOTTI	aree dismesse o da riqualificare		X
		aree di valore storico, ambientale, culturale		X
	SICUREZZA	incidenti		X
	CONGESTIONE	collegamenti congestionati	X	
	INTERFERENZE	soste alle intersezioni		X
Traffico				

Figura 2.18: Indicatori aggregati e disaggregati per zone

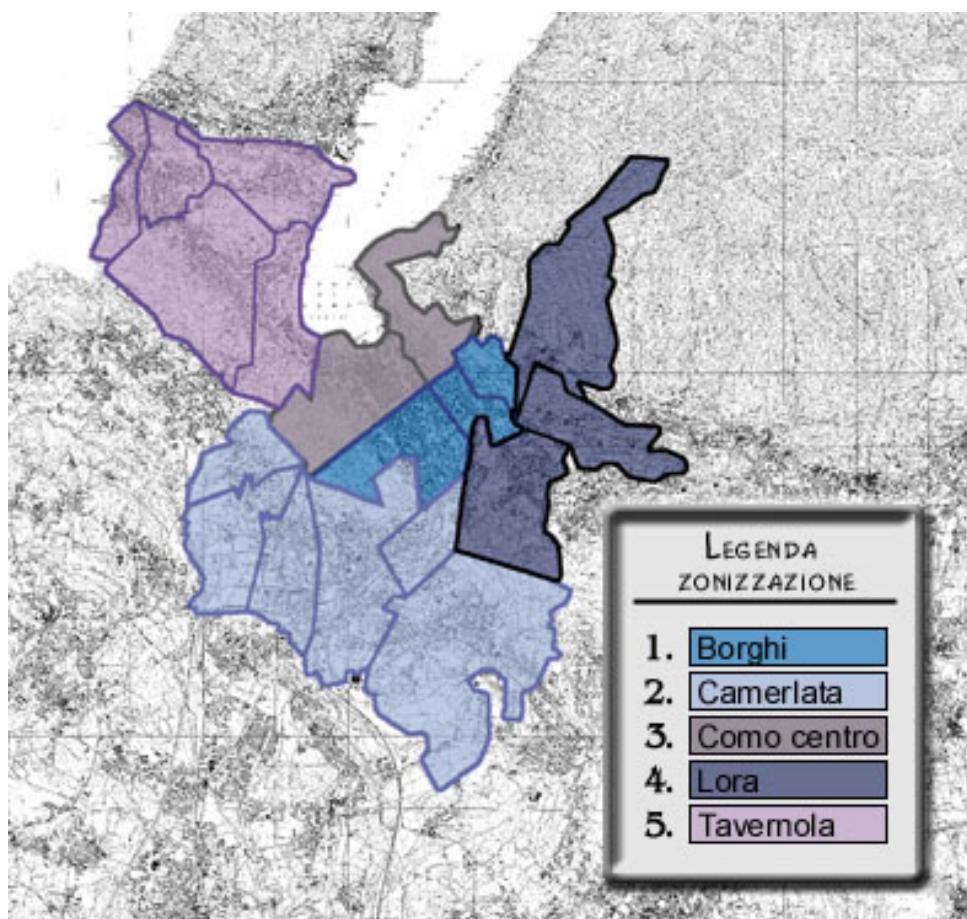


Figura 2.19: Zonizzazione del comune di Como

che possa intervenire a difenderli in modo più o meno favorevole al progetto stesso. L'aspetto fondamentale nella descrizione degli attori è caratterizzare la loro *struttura preferenziale*. Discuteremo in seguito in che cosa questa consista e come possa essere ricavata e descritta. Tuttavia, gli attori potrebbero anche indicare alternative, elementi di scenario e indicatori che erano sfuggiti all'analisi.

Nel caso in esame, si sono individuate quattro categorie, ulteriormente suddivise:

1. Attori istituzionali (Comune, Provincia, Regione)
2. Società
 - Cittadini
 - Associazioni ecologiste
 - Associazioni di categoria
3. Utenti
 - abituali (pendolari, studenti)
 - saltuari
4. Compagnie di trasporto
 - FNM (le ferrovie regionali il cui tracciato fra Grandate e Como Lago dovrebbe ospitare il nuovo servizio);
 - SPT (l'azienda del trasporto pubblico locale che gestisce le linee di autobus, la cui domanda sarebbe sconvolta dal nuovo servizio, richiedendo una revisione radicale dei percorsi e degli orari);
 - FS (le ferrovie statali, la cui stazione di Como S. Giovanni potrebbe ospitare, in parte o integralmente, i treni che ora servono la linea FNM).

Ogni categoria ha interesse e competenze limitate, spesso corrispondenti a uno o pochi sottoalberi dell'albero degli indicatori (questa è un'altra ragione per la modellazione gerarchica dell'impatto). Quindi, verrà coinvolta soltanto (o soprattutto) nelle fasi del processo decisionale che coinvolgono quegli specifici indicatori. La Figura 2.20 descrive una relazione ipotetica fra gli attori e i sottosettori a cui sono verosimilmente interessati.

1. gli attori istituzionali (in particolare il Comune di Como) saranno interessati all'intero sistema degli indicatori;
2. per quanto riguarda la società:
 - i cittadini concentreranno probabilmente il loro interesse sui settori Ambiente, Società e Trasporti;
 - le associazioni ecologiste promuoveranno istanze concentrate sul settore Ambiente;
 - le associazioni di categoria (Camera di Commercio, Industria, Artigianato e Agricoltura - Confesercenti - ...), che promuovono gli interessi economici delle classi produttive, si concentreranno sul settore Economia e sui disagi prodotti dai lavori;
3. gli utenti si concentreranno sul settore Società, e in particolare sulla qualità e l'accessibilità del servizio;
4. per quanto riguarda le compagnie di trasporto:

5. FNM si concentrerà sui settori Società (in particolare Occupazione e Disagi) ed Economia (Costi e Ricavi);
6. SPT è particolarmente interessata ai macrosettori Società, Economia e Trasporti, e in particolare alle ricadute su Occupazione, Disagi, Traffico e i propri Costi e Ricavi;
7. FS è interessata alla gestione del servizio ferroviario eventualmente spostato sulla stazione di Como S. Giovanni.

Sottosettori di interesse per ogni attore

		ATTORI								
Settore	Sottosettore	Comune di Como	Cittadini	Amb./Ecol.	Ass. di categoria	Utenti abituali	Utenti saltuari	FNM	SPT	FS
AMBIENTE	Aria	X	X	X		X	X			
	Rumore	X	X	X		X	X			
	Vibrazioni	X	X	X			X			
	Paesaggio	X	X	X			X			
ECONOMIA	Costi	X						X	X	X
	Ricavi	X						X	X	X
	Valore immobili	X	X		X					
ASPETTI SOCIALE	Apprezzamento	X	X			X	X			
	Disagi	X	X		X	X	X	X	X	X
	Accessibilità	X	X		X	X	X			X
	Occupazione	X			X			X	X	X
TRAFFICO	Sicurezza	X	X	X						
	Congestione	X	X	X		X	X			
	Interferenze	X	X	X		X	X	X	X	

Figura 2.20: Relazione fra gli attori e i sottosettori a cui sono interessati

Individuati gli attori, si deve definire per ciascuno un interlocutore rappresentativo, allo scopo di coinvolgerlo nel processo: definizione del problema, individuazione delle alternative, degli scenari, degli indicatori, calcolo degli impatti, descrizione delle preferenze. Questi interlocutori dovrebbero essere coinvolti via via nelle successive iterazioni durante le quali si ridefiniscono alternative, scenari, impatti, preferenze, ecc... sulla base dei risultati parziali ottenuti durante le iterazioni precedenti dello studio. Gli strumenti di interazione dipendono dalla specifica figura rappresentativa scelta:

1. le interazioni con gli attori istituzionali potrebbero avvenire soprattutto con il sindaco di Como, il governatore della Provincia di Como e quello della Lombardia, o con loro delegati, attraverso riunioni e colloqui approfonditi;
2. riguardo gli attori sociali:
 - le interazioni con i cittadini possono aver luogo con questionari mirati, siti web, incontri pubblici e colloqui con i loro rappresentanti locali (consiglieri di circoscrizione);
 - le interazioni con le associazioni ecologiste e di categoria potrebbero sfruttare questionari, riunioni e colloqui con i loro portavoce;
3. le interazioni con gli utenti potrebbero richiedere sondaggi fra i passeggeri e riunioni con i comitati dei pendolari;
4. le interazioni con le compagnie di trasporto potrebbero aver luogo con dirigenti specificamente incaricati di questo compito.

Si può osservare che ambientalisti, commercianti e utenti sono anche cittadini, oltre che membri di categorie, e quindi esercitano un doppio ruolo. Non è un'inutile

ridondanza? In realtà no, dato che le preferenze dei generici cittadini saranno naturalmente piuttosto diverse da quelle di specifiche categorie, e i mezzi e i modi di esercitarle e manifestarle sono anche diversi. Quindi è corretto rappresentarli come attori diversi, anche se sono in parte sovrapposti. Tenere in considerazione tutti questi aspetti dovrebbe garantire uno svolgimento più fluido all'intero processo di decisione, evitando che ci siano soggetti che si sentono ignorati e decidano di opporsi al progetto in linea di principio.

2.1.6 Il processo di decisione

Per ogni configurazione del sistema, cioè per ogni alternativa e ogni scenario, occorre valutare ogni indicatore. Questo richiederà in alcuni casi di compiere misurazioni fisiche, nella maggior parte dei casi l'applicazione di opportuni modelli descrittivi per calcolare il valore atteso dell'indicatore nella configurazione data. In termini puramente qualitativi:

- le alternative con tecnologia ferroviaria avranno:
 - costi bassi di realizzazione e bassi disagi
 - effetto scarso sull'accessibilità dei punti non serviti dalla linea attuale (ad es., l'ospedale S. Anna)
 - scarsi miglioramenti ambientali
 - quasi nessun impatto occupazionale
 - ...
- le alternative basate su interoperabile avranno:
 - costi economici più alti
 - una maggior penetrazione del servizio nel tessuto cittadino, molto dipendenti anche dal percorso scelto
 - disagi intermedi
 - ...
- le alternative basate su tranvia classica avranno:
 - costi e tempi di realizzazione nettamente più alti
 - un servizio molto capillare
 - ...

L'enorme quantità di valori numerici quantitativi e di valori qualitativi ordinali che ne deriva andrà poi riaggredita riferendosi alle preferenze dei vari attori in gioco. Per ogni attore, bisognerà modellare in qualche modo la struttura preferenziale, in modo da poter stabilire, per una qualsiasi coppia di possibili impatti, quale dei due l'attore considererà preferibile all'altro. Occorre compiere una scelta metodologica di base sul tipo di modello usato per descrivere la struttura preferenziale.

Ciò fatto, occorre aggregare le preferenze dei diversi attori in un ordinamento finale fra le alternative, o quanto meno nella scelta di un'alternativa. Per esempio, alcuni metodi partono dagli indicatori, seguiti dagli scenari, fornendo un ordinamento specifico per ciascun attore; solo alla fine gli ordinamenti dei diversi attori vengono fusi in un ordinamento finale con un processo di negoziazione. Altri metodi prendono in considerazione gli attori fin dal principio. Possibilmente, lo studio dovrebbe anche fornire informazioni sulla *sensitività* di tale ordinamento o scelta, cioè sui limiti entro i quali essa rimane valida, al variare dei dati (scenari, impatti, preferenze, importanza relativa degli attori). Nei capitoli seguenti discuteremo in dettagli questi temi.

2.2 La riapertura dei Navigli a Milano

2.2.1 Il contesto

Benché oggi pochi lo percepiscano, Milano nasce come città d'acque, ricca di fiumi naturali (l'Olona, il Vepra, il Nirone, il Seveso, il Lambro Meridionale e il Lambro), torrenti e rogge, e poi arricchita da un'infinita rete di canali artificiali. La modernità ha condotto alla copertura quasi completa di tale rete per favorire il traffico automobilistico e più di recente alla periodica proposta e discussione di progetti di parziale riapertura. Le Figure 2.21 e 2.22 mostrano la situazione storica e corrente di Milano e della sua rete idrografica.



Figura 2.21: Mappa storica del progressivo allargamento della città di Milano e della sua rete idrografica

Il referendum consultivo del 2011 ha condotto alla stesura di un progetto complessivo, che è disponibile in rete¹⁰

Generalmente, la questione viene impostata e affrontata dal punto di vista urbanistico, storico ed estetico, cioè come *miglioramento del paesaggio e della qualità urbana* e *valorizzazione di un patrimonio culturale* quasi dimenticato, ma ancora molto recente. In realtà, ci sono diverse altre questioni che si intrecciano e che concorrono a rendere potenzialmente interessante l'ipotesi di riapertura. In sintesi, le principali sono:

- la *continuità idraulica*, cioè la possibilità di controllare meglio e ridurre le esondazioni dei corsi d'acqua nella zona nord di Milano, che al momento scorrono intubati sotto la città, con tutti i limiti di portata che questo comporta e la difficoltà di gestire la pulizia dei condotti sotterranei;
- la *navigabilità turistica e commerciale*, con il loro impatto sociale, culturale ed economico (vedi la Figura 2.23 per la situazione attuale);

¹⁰Si vedano gli indirizzi http://www.comune.milano.it/wps/portal/ist/it/servizi/territorio/riapertura_navigli_2016 e <http://www.cuoredimilano.org/index.html>.

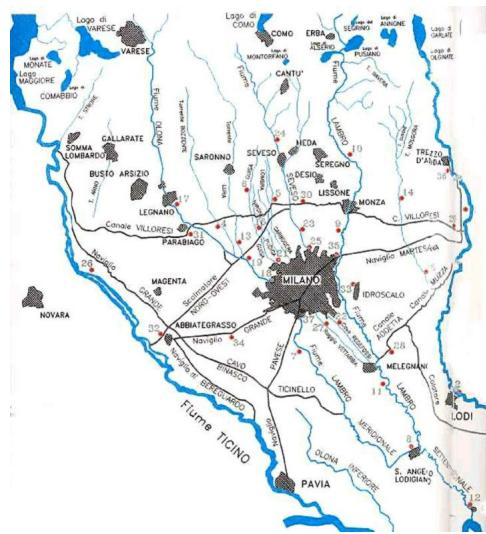


Figura 2.22: Rete idrografica attuale della zona circostante il comune di Milano

- la *produzione di energia elettrica*, non con l'installazione di turbine nei Navigli, ma con l'aumento della portata del Naviglio Pavese in uscita da Milano verso sud, sul quale si trovano diversi piccoli impianti idroelettrici;
 - l'*alimentazione di impianti di riscaldamento a pompa di calore* nella zona della Darsena con uso dell'acqua di falda;
 - il prelievo di acqua attraverso pozzi che ora sono chiusi avrebbe come ulteriore vantaggio un *abbassamento del livello della falda*, che è fortemente salito con la forte deindustrializzazione della città dalla fine degli anni '70, e che attualmente minaccia i condotti della metropolitana e richiede l'uso continuo di pompe in diverse stazioni.

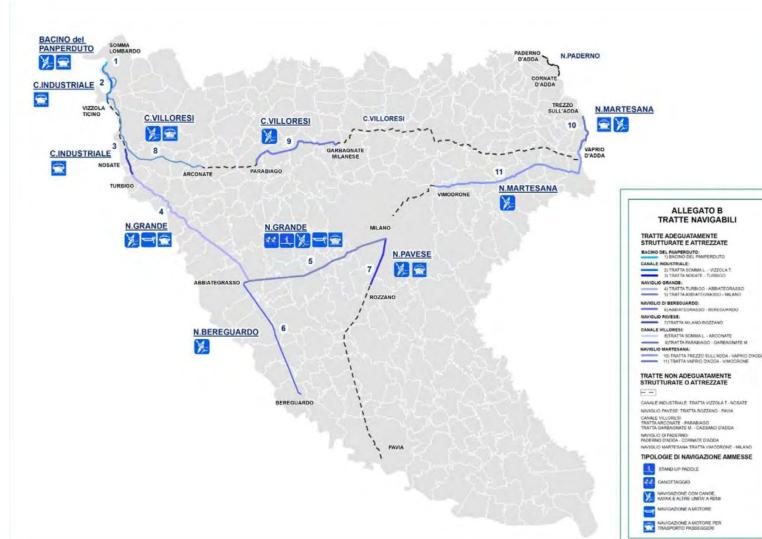


Figura 2.23: Tratte navigabili nell'area intorno a Milano

2.2.2 La definizione delle alternative

Apparentemente, la definizione delle alternative è molto più semplice che per la metrotranvia di Como, dato che i Navigli avevano storicamente un percorso ben definito. In realtà, la cosa è tutt'altro che pacifica. Partendo con una dicotomia di base, si possono distinguere:

- alternative “virtuali”, nelle quali i Navigli vengono resi fruibili al pubblico senza riaprirli veramente, ma con ausilii visivi (cartelloni informativi, fotografie d’epoca, pavimentazioni, ecc. . .); si vedano le Figure 2.24 e 2.25 per alcuni esempi che sono stati proposti come primo passo nel progetto dettagliato che stiamo descrivendo;
- alternative “fisiche”, nelle quali i Navigli vengono effettivamente riaperti.



Figura 2.24: Un primo esempio di pavimentazione, a cubetti di porfido, per far rivivere il percorso storico dei Navigli



Figura 2.25: Un secondo esempio di pavimentazione, con onde dipinte, per far rivivere il percorso storico dei Navigli

Si potrebbe pensare che tutte le alternative fisiche adottino lo stesso percorso, e che esso sia fissato dalla storia (vedi Figura 2.26). In realtà, ci sono molte possibili alternative anche da questo punto di vista:

- la riapertura parziale, lasciando coperti alcuni tratti;
- la riapertura della cinta classica nord-est-sud, che include una serie di varianti minori riguardanti sistemazioni di dettaglio presso:

- il parco di Porta Nuova (vedi la Figura 2.27);
- piazza Cavour (vedi vol. 1 pag. 202 o 276 e 279/281 della documentazione del progetto)
- il Naviglio del Vallone (vedi vol. 1 pagg. 337 e segg. della documentazione del progetto)
- la conca di Viarennna (vedi vol. 1 pag. 206, 342, 344-350 della documentazione del progetto)
- la riapertura dell'intera cerchia interna, anche sul lato ovest: vi sono diversi motivi tecnici che la rendono poco praticabile, primo fra tutti l'interferenza della linea 2 della metropolitana presso Sant'Ambrogio con il percorso originale del Naviglio;
- aperture aggiuntive rispetto al percorso storico:
 - la roggia Vettabbia (vedi vol. 2 pag. 249 della documentazione del progetto)
 - il progetto “Darsena 2”, che prevede un bacino completamente nuovo presso la stazione ferroviaria di Porta Genova (vedi vol. 2 pag. 254 della documentazione del progetto)

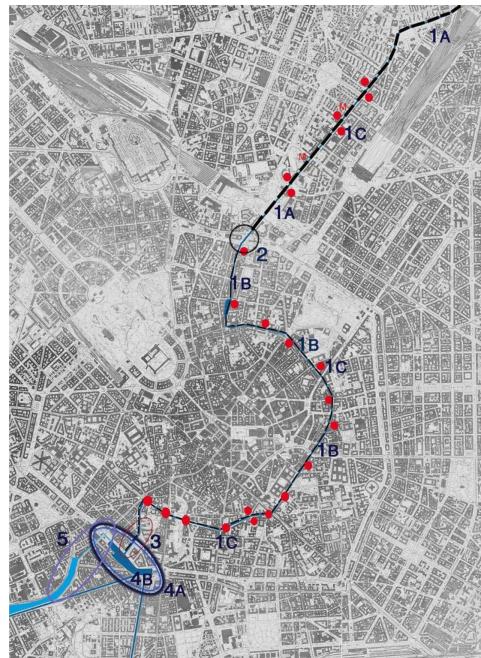


Figura 2.26: Il percorso storico dei Navigli a Milano

Un altro fondamentale elemento di alternativa è la navigabilità: il nuovo canale potrebbe essere aperto al movimento di natanti oppure chiuso. Gli elementi di alternativa (percorso, navigabilità ed eventuali altri) vanno poi combinati per costituire le vere e proprie alternative, scartando le combinazioni che non sono possibili. Per esempio, le alternative virtuali, e le alternative di riapertura parziale molto limitata impediscono ovviamente la navigabilità.

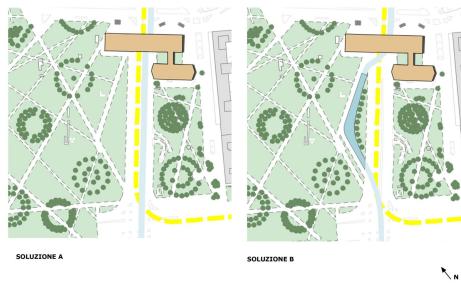


Figura 2.27: Le due soluzioni proposte al termine di via Melchiorre Gioia: A) ipotesi di completo interramento; B) il Naviglio all'aperto nel parco di Porta Nuova

2.2.3 La definizione degli scenari

Questo punto non è sviluppato dal progetto dettagliato che stiamo descrivendo. Ci limitiamo quindi a un'analisi superficiale. I più probabili elementi di scenario sono da ricercare in:

- la disponibilità di finanziamenti pubblici per la realizzazione e la gestione a lungo termine;
- la chiusura del centro al traffico, più o meno rigorosa;
- la realizzazione della linea 4 della metropolitana, che consentirebbe una copertura del centro cittadino a distanza pedonale anche nell'ipotesi che l'accesso al centro col mezzo privato e con i mezzi di superficie diventi più difficile (la Figura 2.28 riporta le aree coperte dalle linee attuali e dalla futura linea 4); inoltre, i lavori della metropolitana, accoppiati a quelli di costruzione del canale sotterraneo di raddoppio dei Navigli per istituire sin da subito la continuità idraulica, consentirebbero una riduzione dei costi¹¹

2.2.4 La definizione e il calcolo degli indicatori

Ancora una volta, si tratta di procedere in maniera gerarchica, individuando prima grandi macrosettori e poi dividendoli in sottosettori via via più specifici:

- l'impatto sull'inquinamento;
- la variazione dei tempi di accesso a vari luoghi della città;
- l'impatto sul traffico privato (vol. 2 pag. 79, 88);
- l'impatto sul trasporto pubblico (vol. 2 pag. 80, 85: velocizzazione), e in particolare con le linee dei tram, che devono usare ponti a livello strada, dato che non possono affrontare pendenze;
- i costi di realizzazione (vol. 1, da pag. 123 a pag. 138);
- l'impatto edonico sui valori immobiliari (vol. 2 pag. 145-146)
- l'impatto sulle imprese commerciali (vol. 2 pag. 150)

¹¹Questo è attualmente previsto dai lavori in corso; c'è chi teme che la costruzione del canale sotterraneo possa rendere meno attraente la riapertura dei Navigli, e quindi sfavorire le alternative "fisiche".

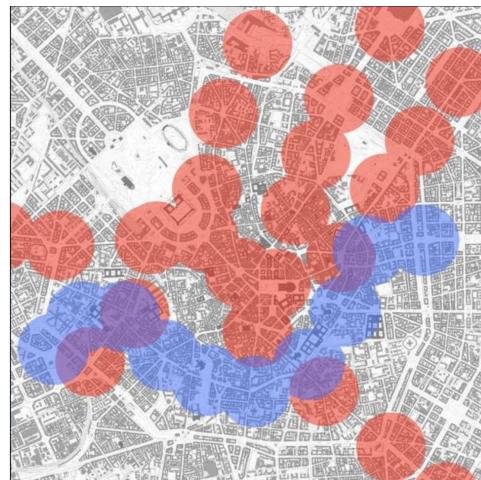


Figura 2.28: Le aree a distanza pedonale dalle stazioni della metropolitana attuali (in rosso) e dalla futura linea 4 (in blu)

Alcuni di questi impatti potrebbero essere così negativi a suggerire un ritorno alla fase di definizione delle alternative per correggerle e adattarle attraverso l'introduzione di misure di riduzione dell'impatto. Per esempio, riaprire i canali colpirebbe il traffico automobilistico lungo il canale e in alcuni punti impedirebbe quasi l'accesso dei veicoli agli edifici circostanti. Questo richiede alcune soluzioni locali per guadagnare spazio, come dei "balconi" sopra il canale (vol. 2 pagg. 292-293-295). E ancora, gli incroci a livello strada con i tram richiedono di tenere il livello dell'acqua sufficientemente basso e di usare natanti appositi (vol. 1 pag. 42) se si vuole garantire anche la navigabilità.

2.2.5 L'organizzazione spaziale e temporale

Diversi indicatori (in particolare quelli sul traffico, l'accessibilità, il valore immobiliare, l'inquinamento, ecc...) si prestano naturalmente ad essere valutati per aree geografiche. Alcuni indicatori vanno quindi riferiti a zone, mentre altri, più strettamente legati al percorso lineare del Naviglio (per esempio, il valore immobiliare) vanno riferiti a tratte (nello studio si pensa a 16 tratte)

Per quanto riguarda le fasi temporali, alcune di queste corrispondono a vere e proprie alternative, nel senso che si può pensare di realizzare una parte dell'opera e poi lasciare in sospeso la parte rimanente a lungo termine o persino a tempo indeterminato nell'attesa che si verifichino le condizioni per completarla. Per esempio, le alternative "virtuali" con pavimentazione e cartelloni si possono anche vedere come una prima fase dell'apertura globale (vol. 2, pag.238). Da un lato, consentono di ottenere un risultato parziale a basso costo e gestire gradualmente il risultato finale; dall'altro, portano a costi complessivamente più alti.

L'ipotesi di raddoppio del Naviglio con una tubazione sotterranea parallela (vol. 2, pag. 239) di per sé ovviamente costituisce un costo aggiuntivo, ma consente anche di ottenere in tempi rapidi il vantaggio della continuità idraulica e di distribuire eventualmente nel tempo i passi successivi, graduando la spesa.

2.2.6 La definizione degli attori

La definizione degli attori ha luogo in modo simile a quanto descritto per la Metrotranvia di Como: si può considerare il sindaco di Milano come decisore, ma si devono tenere in considerazione attori come le società di trasporto (principalmente ATM, che gestisce quella autobus e tram; le linee ferroviarie non sono influenzate dal progetto), le associazioni (ambientaliste e di categoria), i cittadini. Fra questi ultimi, i proprietari di immobili lungo il percorso del canale potrebbero essere trattati a parte.

2.2.7 Il processo di decisione

Il processo di valutazione e decisione dovrebbe seguire le stesse linee adottate per la Metrotranvia di Como. Per prima cosa, si valuta il valore che ciascun indicatore assume in corrispondenza ad ogni configurazione, cioè ogni alternativa e scenario. Questo comporta l'applicazione di opportuni modelli descrittivi. Ci si può aspettare che:

- le alternative “virtuali” siano particolarmente economiche e abbiano impatto trascurabile sul traffico e l'inquinamento, ma non offrano alcun vantaggio dal punto di vista della continuità idraulica (a meno che siano accoppiate alla costruzione di una tubatura sotterranea), della produzione di energia elettrica e della gestione della falda, e un impatto ridotto sul turismo e il commercio; l'impatto estetico è più difficile da valutare;
- le alternative “fisiche” siano nettamente più costose, raggiungano gli obiettivi desiderati per la continuità idraulica, la produzione di energia elettrica e la gestione della falda, favoriscano turismo, commercio e valori immobiliari e abbiano un buon impatto estetico.

E ancora che:

- le alternative che comportano un canale navigabile abbiano vantaggi turistici, commerciali e di accessibilità, ma un costo più alto perché richiedono di gestire un sistema di chiuse per superare il dislivello fra le varie tratte (vedi vol. 2 pag. 8 nella documentazione del progetto);
- le alternative che comportano un canale non navigabile siano meno costose, ma favoriscano meno l'accessibilità, il commercio e il turismo, anche se potrebbero avere un impatto estetico migliore, dato che un canale navigabile richiede che il livello dell'acqua sia abbastanza basso da lasciar passare i natanti sotto i ponti e quindi riduce la visibilità dell'acqua dalle sponde.

Valutati tutti gli indicatori, si deve passare alla fase di aggregazione, nella quale rientra la struttura preferenziale degli attori individuati. Questa fase può avvenire in molti modi, che verranno descritti nei capitoli seguenti, e dovrebbe concludersi con un ordinamento delle alternative, o almeno con la scelta di una di esse, nonché con informazioni sulla sensitività di tale ordinamento o scelta.

Capitolo 3

Definizioni fondamentali e problemi concettuali

Definizione 1 Un problema decisionale è definito dalla sestupla

$$P = (X, \Omega, F, f, D, \Pi)$$

dove

- X è l'insieme delle alternative possibili, o soluzioni ammissibili;
- Ω è l'insieme degli scenari possibili o esiti, anche detto spazio campionario;
- F è l'insieme degli impatti possibili, anche detto spazio degli indicatori;
- $f : X \times \Omega \rightarrow F$ è la funzione impatto, caratterizzata dal fatto che:
 - ogni coppia $(x, \omega) \in X \times \Omega$ descrive una configurazione del sistema sul quale si intende prendere una decisione;
 - la funzione f associa a ogni configurazione (x, ω) del sistema un impatto $f(x, \omega) \in F$;
 - tale impatto $f(x, \omega)$ descrive tutte le caratteristiche della configurazione (x, ω) che sono rilevanti ai fini della decisione (costi, profitti, livelli di qualità, ecc...);
- D è l'insieme dei decisori, anche detti attori;
- $\Pi : D \rightarrow 2^{F \times F}$ è una funzione che associa ad ogni decisore $d \in D$ un sottoinsieme di coppie di impatti, $\Pi(d) \subseteq F \times F$; tale sottoinsieme si interpreta come una relazione binaria e rappresenta la preferenza del decisore d .

Lo scopo del problema è individuare una soluzione $x^* \in X$ o un sottoinsieme di soluzioni $X^* \subseteq X$ che i decisori considerano soddisfacenti sulla base delle loro preferenze fra gli impatti $f(x^*, \omega)$ per $\omega \in \Omega$ e gli altri impatti $f \in F$.

3.1 Alternative

Le alternative descrivono formalmente gli eventi che sono sotto il controllo dei decisorи. Quindi, l'insieme X include tutte le scelte che sono possibili per i decisorи, cioè compatibili con i vincoli fisici, economici, legali, organizzativi, ecc... del sistema.

La parola “alternativa” sottolinea il principio che due diverse soluzioni non possono assolutamente essere applicate entrambe. Se ciò è possibile, si devono definire alternative le loro combinazioni possibili.

In generale, ipotizzeremo $X \subseteq \mathbb{R}^n$, cioè descriveremo quantitativamente le alternative come vettori (per convenzione, colonna) di n numeri reali:

$$x = [x_1, \dots, x_n]^T \quad \text{con } x_i \in \mathbb{R} \text{ per } i = 1, \dots, n$$

Ciascuna componente x_i dell’alternativa x si chiama *elemento di alternativa* o *variabile di decisione* ed è una grandezza elementare che definisce un aspetto dell’alternativa stessa.

L’ipotesi $X \subseteq \mathbb{R}^n$ è in parte limitante perché

- esclude spazi X a dimensione infinita, dove le scelte a disposizione dei decisori sarebbero vere e proprie funzioni dipendenti da un insieme infinito di parametri (ad es., non consideriamo problemi di controllo ottimo).

D’altra parte questa ipotesi:

- ammette problemi le cui soluzioni sono famiglie di funzioni dipendenti da un numero finito di parametri (ad es., si può scegliere fra tutti i polinomi di grado $\leq n - 1$);
- ammette problemi in cui le alternative sono descritte in termini qualitativi: basta assegnare alle variabili x_i dei valori numerici convenzionali in corrispondenza ai gradi della scala qualitativa (ad es., da 1 a 10);
- ammette problemi in cui le alternative sono banalmente enumerate: basta introdurre una variabile di decisione binaria per ogni alternativa e imporre che esattamente una variabile sia pari a 1 (in questo modo, l’alternativa i^* corrisponde al vettore caratteristico x , con $x_{i^*} = 1$ e $x_i = 0$ per ogni $i \neq i^*$).

Si noti infatti che, per ora, la notazione è puramente descrittiva, cioè non abbiamo ancora deciso di fare calcoli con il vettore x .

Possiamo anche fare una classificazione dei problemi decisionali in base alla cardinalità di X , dividendoli in *discreti* o numerabili (cioè in corrispondenza biunivoca con un sottoinsieme di \mathbb{N}) e *continui*. I problemi discreti sono a loro volta divisibili in infiniti e finiti. Per concludere, i problemi finiti si dicono *combinatorici* se la loro cardinalità, benché finita, è esponenziale rispetto al numero di variabili di decisione n . Questo avviene di solito quando le alternative si ottengono combinando un numero finito di valori possibili per ciascuna delle variabili di decisione. Questa classificazione ha influenza sui metodi usati per risolvere i problemi stessi: ogni categoria infatti, avrà i suoi metodi possibili, o quanto meno preferibili.

3.2 Scenari

Gli scenari descrivono formalmente gli eventi fuori dal controllo dei decisori, ma con una ricaduta significativa sul sistema. Quindi, l’insieme Ω definisce tutti gli eventi incontrollabili che sono possibili nel sistema durante l’orizzonte temporale che interessa la decisione e che impattano su di essa.

In generale, ipotizzeremo $\Omega \subseteq \mathbb{R}^r$, cioè descriveremo gli scenari quantitativamente come vettori (colonna) di r numeri reali.

$$\omega = [\omega_1, \dots, \omega_r]^T \quad \text{con } \omega_k \in \mathbb{R} \text{ per } j = 1, \dots, r$$

Ciascuna componente ω_k dello scenario ω si chiama *elemento di scenario* o *variabile esogena* ed è una grandezza elementare che definisce un aspetto dello scenario stesso.

Vale quanto detto sopra per le alternative: questa ipotesi esclude scenari a dimensioni infinite, ma ammette scenari descritti attraverso funzioni parametriche, scale qualitative o semplicemente enumerati.

Gli scenari incorporano qualsiasi fonte di incertezza sul comportamento del sistema:

- fenomeni veramente esogeni che influenzano il sistema e sarebbero forse prevedibili se allargassimo i confini del sistema stesso (ad es., lanciando una moneta, se conoscessimo alla perfezione l'andamento del vento, potremmo forse prevedere l'esito del lancio; l'andamento del vento però è esterno al sistema moneta, per cui l'esito del lancio ricade nello scenario);
- effetti di una modellazione incompleta o parzialmente errata del sistema; in tal caso, si parla spesso di *disturbi* (ad es., gestendo una valvola idraulica, piccole perdite possono far passare dell'acqua anche quando la valvola è chiusa; queste perdite non sono esogene, ma vengono incorporate nello scenario per consentire un modello semplificato di valvola perfetta).

3.3 Impatti e funzione impatto

Gli impatti descrivono formalmente tutto ciò che è rilevante ai fini della decisione. Ad esempio, se vogliamo regolare la temperatura di una stanza, possiamo decidere la posizione del termostato (che ricade nelle alternative) e sappiamo che essa dipenderà anche dalla temperatura esterna (che ricade nello scenario), ma ciò che interessa ai fini della decisione è la temperatura interna della stanza. Questa costituisce l'impatto, e dipende sia dalla posizione del termostato sia dalla temperatura esterna.

Poiché la decisione viene presa sulla base dell'impatto, configurazioni (x, ω) e (x', ω') con impatto identico ($f(x, \omega) = f(x', \omega')$) sono assolutamente indistinguibili tra loro. È quindi fondamentale che il modello espliciti tutti gli elementi che compongono l'impatto, per evitare di falsare la relazione di preferenza. Questo sembra ovvio, ma in pratica succede di frequente che soluzioni giudicate soddisfacenti e realizzate in pratica, si rivelino poi mediocri perché alcuni elementi dell'impatto non erano stati considerati nel modello. È il caso, ad esempio, delle opere pubbliche bloccate da interventi legali oppure ostacolate dall'opinione pubblica: chi aveva modellato il processo decisionale non aveva considerato aspetti come la reazione dei cittadini, che può far parte integrante dell'impatto. Da un lato, aggiungere elementi al modello lo rende più complicato e costoso da costruire (bisogna procurarsi i dati o mettersi in grado di prevederli) e da risolvere (le dimensioni del modello crescono). Dall'altro, per'ò, un modello scorretto può portare a una decisione sbagliata, e costosa.

In generale, ipotizzeremo $F \subseteq \mathbb{R}^p$, cioè descriveremo gli impatti quantitativamente come vettori (colonna) di p numeri reali.

$$f = [f_1, \dots, f_p]^T \quad \text{con } f_l \in \mathbb{R} \text{ per } l = 1, \dots, p$$

con la solita avvertenza sugli impatti a dimensione infinita (esclusi da questa definizione), e sugli impatti qualitativi o enumerati (compresi in essa).

Ciascuna componente $f_l \in \mathbb{R}$ dell'impatto f si chiama *indicatore*, o anche *criterio, attributo, obiettivo* (in questi ultimi casi, si assume di solito che rappresentino dei benefici o dei costi, e che quindi i decisori preferiscano impatti con valori più alti

o più bassi). Ogni indicatore è una grandezza elementare che definisce un aspetto dell’impatto stesso. Dovrebbe essere comprensibile a ogni decisore, univoco (cioè non mescolare aspetti diversi), non ridondante (cioè non ripetere quanto già espresso da altri indicatori). Ogni indicatore dipende in modo calcolabile dall’alternativa scelta e dallo scenario che si realizza attraverso la funzione $f_i = f_i(x, \omega)$.

Contrariamente a tutte le altre definizioni presentate sinora, il termine “impatto” non è di uso comune in letteratura, ma è stato coniato appositamente per questo corso, in modo da poter distinguere i singoli indicatori f_i dalla loro combinazione f , là dove questo si riveli utile.

Rappresentazioni dell’impatto nel caso finito

Nel caso finito si può rappresentare la funzione impatto f con una *matrice di valutazione*, in cui le righe sono associate alle configurazioni (x, ω) e le colonne agli indicatori f_i : ogni cella della matrice riporta il valore dell’indicatore di colonna nella configurazione di riga (vedi Tabella 7.1).

$f(x, \omega)$	f_1	f_2	f_3	f_4	f_5	f_6
(x_1, ω_1)	10	5	40	20	24	180
(x_1, ω_2)	16	10	60	16	20	190
(x_2, ω_1)	20	6	23	8	17	230
(x_2, ω_2)	24	8	50	12	10	100

Tabella 3.1: Rappresentazione della funzione impatto per un problema finito con due alternative, due scenari e sei indicatori attraverso la matrice di valutazione

Sempre nel caso finito, la funzione impatto f si può anche rappresentare con una *rosa dei venti*, cioè un grafico bidimensionale che ha un asse per ogni indicatore f_i e rappresenta una configurazione (x, ω) con un poligono a p vertici che unisce i punti degli assi che corrispondono ai valori degli indicatori nella configurazione stessa (vedi Figura 3.1). Si noti come ogni indicatore abbia la propria unità di misura, indipendente dalle altre.

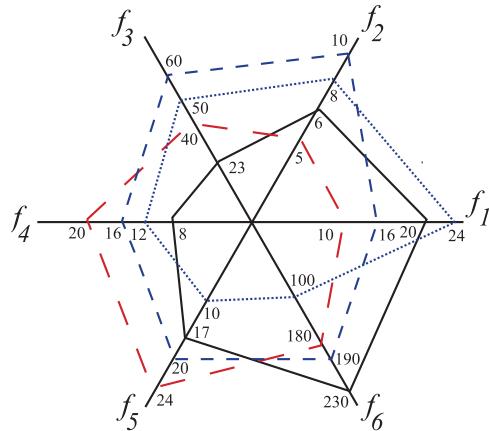


Figura 3.1: Rappresentazione della funzione impatto per un problema finito con due alternative, due scenari e sei indicatori attraverso una rosa dei venti

3.4 Decisori

Decisori sono tutti coloro che contribuiscono alla decisione. Nel caso di soggetti che non contribuiscono direttamente alla decisione, ma esercitano comunque un ruolo con le loro preferenze, si parla spesso di attori.

L'insieme dei decisori si intende come un semplice insieme finito D , senza ulteriori caratterizzazioni. In particolare, non è necessario supporre che sia immerso in uno spazio euclideo come gli altri insiemi introdotti sinora.

Un aspetto che non appare esplicitamente dalla Definizione 1 è la relazione fra decisori e variabili di decisione. In effetti, questa relazione cambia a seconda del problema:

- in alcuni problemi, ogni decisore ha il completo controllo su un sottoinsieme delle variabili di decisione disgiunto da quelli controllati dagli altri;
- in altri casi, tutti i decisori controllano tutte le variabili di decisione, e devono coordinarsi prima di definire i valori di queste ultime.

Ovviamente, si può pensare a casi intermedi, anche se la loro trattazione in letteratura è estremamente rara (se mai sono stati trattati). Dal punto di vista della notazione descrittiva, questa distinzione non ha molta importanza:

- nel caso di variabili gestite indipendentemente, in genere l'insieme delle soluzioni ammissibili si ottiene combinando con un prodotto cartesiano gli insiemi ammissibili per i singoli decisori: $X = X^{(1)} \times \dots \times X^{(d)} \times \dots X^{(|D|)}$.
- nel caso di variabili gestite in comune, in genere l'insieme delle soluzioni ammissibili si ottiene intersecando quelli dei singoli decisori: $X = X^{(1)} \cap \dots \cap X^{(d)} \cap \dots X^{(|D|)}$.

In entrambi i casi, scenari e impatti funzionano come già descritto, e le preferenze funzionano come descritto in seguito, senza differenze particolari. Saranno invece molto diversi i metodi per affrontare e risolvere questi problemi.

3.5 Relazione di preferenza

Per un decisore, *preferire l'impatto* $f \in F$ all'impatto $f' \in F$ significa *considerare accettabile la sostituzione di* f' *con* f , cioè del secondo impatto con il primo. Questo corrisponde a stabilire una relazione orientata fra coppie di impatti. Infatti, la funzione $\Pi : D \rightarrow 2^{F \times F}$ associa ad ogni decisore d un sottoinsieme $\Pi(d)$ di coppie ordinate di impatti, sottoinsieme che da ora in avanti chiameremo Π_d . Il significato di questa funzione è che:

- per ogni coppia $(f, f') \in \Pi_d$ il decisore d preferisce f a f' , cioè accetta di scambiare il secondo impatto f' con il primo impatto f ;
- l'elenco di tutte le coppie di P_d descrive completamente le preferenze del decisore d .

L'insieme Π_d si può interpretare come una *relazione binaria* sull'insieme impatto F :

$$\Pi_d = \{(f, f') \in F \times F : d \text{ preferisce } f \text{ a } f'\}$$

Nel confronto fra due impatti f e f' , se il decisore d preferisce f rispetto a f' , al posto di scrivere $(f, f') \in \Pi_d$, si è soliti usare la notazione

$$f \preceq_d f'$$

Nel seguito, ogni volta che vi sia un solo decisore oppure che si consideri un decisore generico, e quindi ogni volta che si consideri una sola relazione di preferenza anziché una famiglia dipendente da D , sopprimeremo per semplicità l'indice d , scrivendo Π e \preceq al posto di Π_d e \preceq_d .¹

Inoltre, data una relazione binaria Π , definiamo

- *relazione complemento* $\bar{\Pi}$ la relazione che lega tutte e sole le coppie non legate da Π : $\bar{\Pi} = (F \times F) \setminus \Pi$ e anziché $(f_1, f_2) \in \bar{\Pi}$ si usa scrivere $f_1 \not\preceq f_2$
- *relazione inversa* Π^{-1} la relazione che inverte tutte e sole le coppie legate da Π : $\Pi^{-1} = \{(f_1, f_2) \in F \times F : (f_2, f_1) \in \Pi\}$; anziché $(f_1, f_2) \in \Pi^{-1}$ si usa scrivere $f_1 \succeq f_2$

3.5.1 Rappresentazioni della preferenza nel caso finito

Se l'insieme degli impatti F è finito, si può rappresentare una generica relazione di preferenza Π con una *matrice di incidenza* di $|F|$ righe e colonne (una per ciascun impatto), il cui elemento (f, f') sia pari a 1 quando $f \preceq f'$, a 0 altrimenti. La Tabella 3.2 rappresenta una relazione di preferenza definita su uno spazio F con cinque impatti.

	f	f'	f''	f'''	f''''
f	1	0	1	1	1
f'	1	1	1	1	1
f''	0	0	1	1	1
f'''	0	0	0	1	0
f''''	0	0	1	0	1

Tabella 3.2: Matrice di incidenza per una relazione di preferenza Π su uno spazio con cinque impatti

Alternativamente, una relazione di preferenza su uno spazio finito si può rappresentare con un grafo orientato $G = (F, \Pi)$, i cui nodi corrispondono agli impatti, mentre gli archi corrispondono alle coppie che appartengono alla relazione stessa. La Figura 3.2 riporta il grafo associato alla stessa relazione di preferenza rappresentata nella Tabella 3.2.

3.5.2 Relazioni derivate

Data una relazione di preferenza Π , si possono derivare altre tre relazioni con un significato interessante:

1. l'*indifferenza* $\text{Ind}_{\Pi} = \Pi \cap \Pi^{-1} = \{(f, f') \in F \times F : f \preceq f', f' \preceq f\}$ raccoglie le coppie di impatti che il decisore accetta di scambiare in entrambi i versi; quando $(f, f') \in \text{Ind}_{\Pi}$, più semplicemente scriveremo che $f \sim f'$;

¹Indicando con Π una generica relazione $\Pi_d \subseteq F \times F$ non si faccia confusione con la funzione $\Pi : D \rightarrow 2^{F \times F}$. Quale dei due significati attribuire al simbolo Π dovrebbe essere chiaro dal contesto. L'uso del medesimo simbolo è giustificato dal fatto che abbiamo definito $\Pi_d = \Pi(d)$.

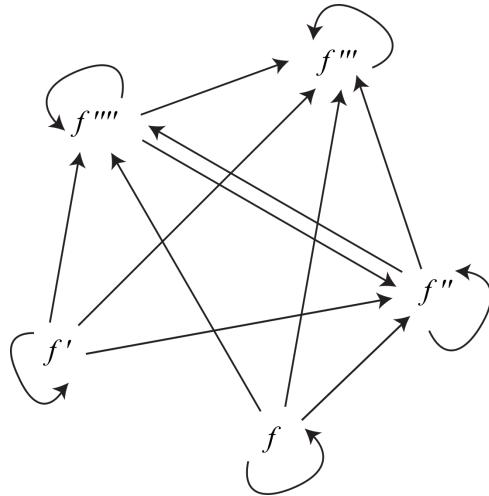


Figura 3.2: Grafo associato ad una relazione di preferenza Π su uno spazio con cinque impatti

2. la *preferenza stretta*: $\text{Str}_{\Pi} = \Pi \cap \bar{\Pi}^{-1} = \Pi \setminus \Pi^{-1} = \{(f, f') \in F \times F : f \preceq f', f' \not\preceq f\}$ raccoglie le coppie di impatti che il decisore accetta di scambiare solo nel verso indicato; quando $(f, f') \in \text{Str}_{\Pi}$, più semplicemente scriveremo che $f \prec f'$;
3. l'*incomparabilità* $\text{Inc}_{\Pi} = \bar{\Pi} \cap \bar{\Pi}^{-1} = \{(f, f') \in F \times F : f \not\preceq f', f' \not\preceq f\}$; raccoglie le coppie di impatti che il decisore non accetta di scambiare in alcun verso; quando $(f, f') \in \text{Inc}_{\Pi}$, più semplicemente scriveremo che $f \bowtie f'$.

Che cosa rappresentino indifferenza e preferenza stretta è abbastanza ovvio. L'incomparabilità modella le situazioni in cui il decisore non è indifferente fra due impatti, ma è incapace o non disposto a scegliere fra loro: se gli si propone uno scambio lo rifiuta, ma rifiuta anche lo scambio opposto. Esempi tipici in letteratura sono quelli della scelta fra uccidere un animale e distruggere un capolavoro d'arte oppure fra uccidere una persona o un'altra.

Nel caso della relazione rappresentata nella Figura 3.2, l'indifferenza lega gli impatti f'' e f'''' e ogni impatto a se stesso, la preferenza stretta lega le altre coppie fra le quali esisteva una preferenza nella relazione originale, l'incomparabilità lega gli impatti f e f' . Le tre relazioni derivate sono rappresentate nelle tre parti della Figura 3.3.

Per ogni coppia di impatti (f, f') , possono quindi valere quattro casi alternativi:

1. $f \prec f'$: f è strettamente preferibile a f' ;
2. $f' \prec f$: f' è strettamente preferibile a f ;
3. $f \sim f'$: f e f' sono indifferenti;
4. $f \bowtie f'$: f e f' sono incomparabili.

Quindi, la preferenza stretta, l'inverso della preferenza stretta, l'indifferenza e l'incomparabilità formano una partizione di $F \times F$. Analogamente, preferenza stretta e indifferenza formano una partizione della preferenza originale. Questo si può facilmente verificare confrontando le tre parti della Figura 3.2 con la Figura 3.3.

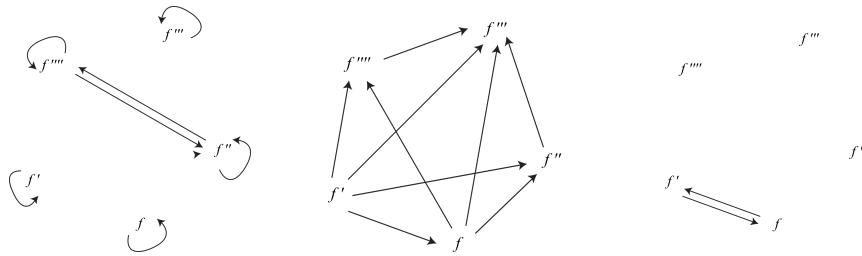


Figura 3.3: Grafi associati alle relazioni di indifferenza (a sinistra), preferenza stretta (al centro) e incomparabilità derivate dalla relazione di preferenza della Figura 3.2

Merita un accenno il fatto che diversi testi adottano un’assiomatica alternativa, ma equivalente a quella appena presentata. Invece di partire da una relazione intepretata come preferenza e derivarne altre tre, si può partire da una coppia di relazioni (sottoinsiemi di $F \times F$) fra loro disgiunte, interpretarle come preferenza stretta e indifferenza, e ricavarne le due relazioni mancanti: la preferenza (che in questo caso si definisce *debole*) sarà l’unione di preferenza stretta e indifferenza, mentre l’incomparabilità sarà l’insieme delle coppie che non appartengono né alla preferenza né al suo inverso.

3.5.3 Proprietà della relazione di preferenza

Fino a questo punto, una relazione di preferenza è una generica relazione binaria su F , cioè una generica collezione di sottoinsiemi estratti da $F \times F$. Non tutte le relazioni binarie, però, esprimono relazioni di preferenza sensate per un decisore. Idealmente, si vorrebbe che la relazione di preferenza fosse:

- *realistica*, cioè modellasse correttamente le intenzioni del decisore;
- *efficace*, cioè consentisse algoritmi che portino a una scelta soddisfacente.

Abbastanza sorprendentemente, le due esigenze di realisticità ed efficacia tendono spesso a configgere.

Ciò non toglie che entrambe le esigenze richiedano opportune proprietà aggiuntive alla relazione di preferenza, oltre al fatto banale di essere binaria. Esiste una certa libertà nella scelta di quali proprietà imporre, e molte di loro pongono problemi concettuali di grande rilevanza. In ultima analisi, la scelta del modello di preferenza va motivata sulla base degli scopi del modello stesso. Cominciamo a introdurre alcune possibili proprietà e poi discutiamo dei problemi concettuali o tecnici che esse comportano.

Una relazione binaria Π su un insieme F è:

- *riflessiva* quando $I \subseteq \Pi$, dove I è la relazione di identità, ovvero $f \preceq f$ per ogni $f \in F$: ogni impatto è preferibile a se stesso (debolmente, cioè in realtà è indifferente);
- *antisimmetrica* quando $\Pi \cap \Pi^{-1} \subseteq I$, ovvero $f \preceq f'$ e $f' \preceq f \Rightarrow f = f'$ per ogni $f, f' \in F$: due impatti sono indifferenti fra loro solo quando sono esattamente identici;

- *completa* quando $\Pi \cup \Pi^{-1} = F \times F$, ovvero $f \not\preceq f' \Rightarrow f' \preceq f$ per ogni $f, f' \in F$: dati due impatti, uno dei due è sicuramente preferibile all'altro (al limite, sono indifferenti);
- *transitiva* quando $f \preceq f'$ e $f' \preceq f'' \Rightarrow f \preceq f''$ per ogni $f, f', f'' \in F$: se un impatto è preferibile a un altro e questo a un terzo, il primo è preferibile al terzo;
- *simmetrica* quando $f \preceq_{\Pi} f' \Rightarrow f' \preceq f$ per ogni $f, f' \in F$; in generale una relazione di preferenza non gode di questa proprietà (anzi, la preferenza forte è asimmetrica: $f \prec f' \Rightarrow f' \not\prec f$ per ogni $f, f' \in F$), ma la menzioniamo perché ne godono le relazioni derivate di indifferenza e di incomparabilità ne godono, per la loro stessa definizione.

Riflessività In genere si considera pacifico che una relazione di preferenza debba essere riflessiva: non avrebbe senso considerare un decisore incapace di paragonare un impatto con se stesso. Di conseguenza, l'indifferenza è sempre riflessiva, mentre l'incomparabilità e la preferenza stretta non lo sono mai, cioè sono irriflessive.

Antisimmetria L'antisimmetria è invece una condizione piuttosto forte: impone che si possa essere indifferenti solo nel caso di impatti perfettamente identici. Il modello di preferenza proposto da Pareto accetta questa ipotesi (vedi Capitolo 5), mentre altri la rifiutano.

Completezza La completezza è anch'essa una condizione piuttosto forte: esclude l'incomparabilità, cioè richiede al decisore di avere sempre una preferenza fra due impatti (eventualmente, indifferenza). Il decisore è quindi in grado di mettere in fila gli impatti dal migliore al peggiore, eventualmente con degli ex-aequo. Un esempio tipico sono le classifiche dei campionati sportivi. I modelli di preferenza basati sull'uso di funzioni obiettivo accettano questa ipotesi, mentre altri la rifiutano.

Transitività Se la preferenza è transitiva, lo sono anche per definizione le relazioni derivate di preferenza stretta e indifferenza. La transitività è la proprietà più delicata e discussa in assoluto: alcuni la considerano imprescindibile, tanto da vederla come un sinonimo di *razionalità*, mentre altri la contestano. Diversi autori cercano di mostrare con esperimenti mentali la tesi che non si possano risolvere problemi decisionali in maniera sensata senza richiedere preferenze transitive. Gli autori critici cercano di smontare le implicazioni tratte da tali esperimenti e di mostrare sperimentalmente che le preferenze espresse nel mondo reale spesso non sono transitive. Discuteremo questo punto in dettaglio nel seguito.

3.5.4 Relazioni di preferenza più comuni

Le relazioni di preferenza usate più comunemente sono:

- relazioni di *preordine*: godono delle proprietà *riflessiva* e *transitiva*;
- relazioni di *ordine parziale*: godono delle proprietà *riflessiva*, *transitiva* e *antisimmetrica*;
- relazioni di *ordine debole*: godono delle proprietà *riflessiva*, *transitiva* e di *completezza*;

- relazioni di *ordine totale*: godono delle proprietà *riflessiva*, *transitiva*, *antisimmetrica* e di *completezza*.

Gli ordini totali combinano le proprietà di quelli parziali e deboli, e quindi in essi il decisore è in grado di mettere in fila gli impatti dal migliore al peggiore senza ex-aequo. Un esempio tipico sono i numeri reali.

Per quanto riguarda le relazioni derivate, qualora la preferenza sia riflessiva e transitiva, l'indifferenza risulta riflessiva, transitiva e simmetrica, ed è quindi una *relazione di equivalenza*, cioè partiziona l'insieme degli impatti F in classi, ognuna delle quali contiene impatti tutti fra loro indifferenti.

3.5.5 Problemi critici sull'ipotesi di transitività

Merita una discussione approfondita l'ipotesi di transitività. La transitività è la condizione richiesta più spesso a una relazione di preferenza, spesso interpretata come un sinonimo di razionalità. Secondo i sostenitori di questa tesi, negare la transitività porterebbe a effetti paradossali.

Per prima cosa, mostriamo con un esempio che cosa voglia dire negare la transitività. Supponiamo che un decisore, posto di fronte alla scelta fra A e B (un'albicocca e una banana, ad esempio), affermi di preferire A e, posto di fronte alla scelta fra B e C (una banana e una noce di cocco, ad esempio), affermi di preferire B . Un decisore transitivo dovrebbe necessariamente preferire A a C . Un ipotetico decisore non transitivo, invece, potrebbe affermare di non saper scegliere tra A e C (l'albicocca e la noce di cocco), se non addirittura di preferire strettamente C ad A . Consideriamo i due casi e le loro conseguenze paradossali.

1) Incomparabilità contraddette da catene di scambi Se il decisore afferma di preferire A a B e B a C , ma di non saper scegliere tra A e C , si può facilmente indurlo a cedere C (l'arancia) in cambio di A (la mela) attraverso una catena di due scambi che usa come passaggio intermedio proprio B (la pera). Questo suggerisce che, in realtà, il decisore sia in grado di scegliere fra le due opzioni, nonostante abbia affermato di non esserlo. Secondo i proponenti, questo esperimento mentale dimostrerebbe che questo genere di intransitività è poco convincente.

2) Money pump Se invece il decisore afferma di preferire A a B e B a C , e di preferire strettamente C ad A , si potrebbe portarlo ad accettare una catena ciclica di scambi spillandogli anche del denaro. Gli si darebbe B al posto di A , poi C al posto di B e infine, sfruttando la preferenza stretta, A al posto di C più una somma di denaro sufficientemente piccola da non ribaltare la preferenza. A questo punto, il decisore sarebbe tornato al punto di partenza (l'arancia A) avendo perso denaro per niente. Questa è quella che gli economisti chiamano *money pump*. Questo esperimento mentale presenta diversi punti da analizzare a fondo. Infatti, introduce il concetto di denaro, cioè complica l'impatto a comprendere un elemento, o indicatore, aggiuntivo. Inoltre, assume che il denaro sia frazionabile in quantità abbastanza piccole da garantire che l'aggiunta di una tale quantità di denaro a uno degli impatti confrontati non modifichi il segno della preferenza stretta fra loro. Sono tutte ipotesi da valutare attentamente.

I sostenitori della possibilità di usare preferenze non transitive propongono da un lato esperimenti mentali alternativi per dar loro un'interpretazione ragionevole, dall'altro esperimenti pratici che mostrano come gli esseri umani si comportino in modo non transitivo.

1) L'influenza del tempo Le diverse scelte che costituiscono la relazione di preferenza vengono sempre intese come atemporali, ovvero fissate una volta per tutte. In realtà, una decisione spesso consiste di più scelte elementari eseguite in istanti diversi; è possibile che la preferenza cambi nel tempo, dando luogo a violazioni della transitività. Ad esempio, magari oggi il decisore preferisce A a B , domani B a C e dopodomani preferisce C ad A . Questo rende il ciclo della *money pump* non solo possibile, ma verosimile, tanto da stare alla base di attività finanziarie. Queste violazioni della transitività si potrebbero sistemare in un modello transitivo assegnando agli impatti anche un indice temporale, e definendo la relazione di preferenza sugli impatti estesi, cioè sulle coppie impatto-tempo. Questo però complica moltissimo il modello della preferenza. Siccome lo scopo del modello è aiutare a prendere decisioni, usare un modello molto complesso non è in genere desiderabile.

2) Capacità discriminatoria limitata Il decisore non è effettivamente in grado di distinguere perfettamente tra loro gli impatti. Esistono impatti estremamente simili fra i quali il decisore è indifferente perché incapace di distinguerli. Tuttavia, una catena di tali impatti può legare fra loro impatti molto diversi. Ad esempio, ammettiamo che il decisore preferisca il caffè amaro. Posto di fronte a una tazzina con n granelli di zucchero e ad una con $n+1$ granelli di zucchero, il decisore potrebbe però essere incapace di distinguerle, e quindi considerarle indifferenti. Attraverso una catena di tazzine con un numero lentamente crescente di granelli di zucchero, si può arrivare a dimostrare per transitività che egli è indifferente fra un caffè molto amaro e uno molto dolce. In breve, o si assume una capacità infinita di discriminazione (che è illusoria) oppure si rinuncia alla transitività della preferenza.

3) Gli effetti di contesto, o *framing* Il modo in cui gli impatti sono presentati influisce pesantemente sulle preferenze, anche nel caso di impatti del tutto identici.

In un celebre esperimento di psicologia, si chiese a un campione di persone di immedesimarsi nel ministro della sanità di un paese in cui è scoppiata un'epidemia e valutare gli impatti di tre possibili alternative:

1. non fare nulla, lasciando morire 600 persone (impatto f);
2. realizzare il programma A , salvando tutti con probabilità $1/3$ e nessuno con probabilità $2/3$ (impatto f');
3. realizzare il programma B , salvando 200 persone (impatto f'').

La maggioranza del campione scelse il programma B , mostrando di preferire il relativo impatto. È vero che statisticamente l'impatto atteso è lo stesso del programma A , ma affidarsi al gioco delle probabilità non è la stessa cosa che puntare su una certezza, per cui è ragionevole che alcuni decisori possano avere una preferenza stretta del tipo $f'' \prec f'^2$. Si chiese quindi a un altro campione di decisori, considerato equivalente al primo, di valutare gli impatti delle tre seguenti alternative:

1. non fare nulla, lasciando morire 600 persone (impatto f);
2. realizzare il programma A , salvando tutti con probabilità $1/3$ e nessuno con probabilità $2/3$ (impatto f');
3. realizzare il programma B , lasciando morire 400 persone (impatto f'').

²Anche la preferenza opposta è ragionevole, solo è meno comune. Ne ripareremo affrontando le decisioni in condizioni di rischio. Ora l'aspetto probabilistico è irrilevante per l'esperimento e serve solo a occultarne l'aspetto fondamentale

Non è chi non veda che le tre alternative proposte ai due campioni sono identiche, ma che f e f' sono descritte con le stesse parole, mentre f'' è descritta con parole diverse (“lasciar morire” invece che “salvare”). La maggioranza del secondo campione scelse il programma A , cioè espresse la preferenza $f' \prec f''$. Il contesto, cioè l’idea di “salvare”, piuttosto che di “lasciar morire”, persone con certezza cambia la percezione degli impatti.

Modelli della preferenza: descrittivi o prescrittivi?

A questo punto, possiamo tirare qualche somma sul problema di scegliere un modello per la preferenza dei decisori. La scelta della relazione di preferenza dipende in larga misura dallo scopo del modello stesso. L’esigenza di avere una descrizione realistica delle preferenze del decisore spinge verso l’idea di un modello descrittivo fedele, mentre l’esigenza di avere una descrizione efficace per arrivare a una scelta utile spinge verso l’idea di un modello prescrittivo, che vincola la libertà del decisore per “insegnargli” a prendere una decisione buona. Secondo i casi, sarà preferibile privilegiare uno o l’altro approccio, ma in generale i due estremi sono entrambi pericolosi.

3.5.6 Ordini deboli e funzioni valore

Si è già accennato al fatto che gli ordini deboli consentono di stabilire una classifica fra gli impatti, con possibili ex-aequo, come se ad ogni impatto fosse associato un punteggio. Se esistesse effettivamente un punteggio per ogni impatto, la scelta di un impatto buono corrisponderebbe alla scelta del punteggio più alto o più basso, cioè alla massimizzazione di una funzione di utilità oppure alla minimizzazione di una funzione di costo. Questo aprirebbe la strada all’uso degli algoritmi di ottimizzazione per risolvere i problemi decisionali.

Per prima cosa verifichiamo che l’esistenza di un punteggio porta effettivamente a godere delle proprietà degli ordini deboli. Quindi, ci chiederemo se valga anche il contrario, cioè se qualsiasi decisore abbia come relazione di preferenza un ordine debole si possa trattare come se attribuisse un punteggio a ogni impatto. Vedremo che le cose sono un po’ più complicate, ma non molto lontane da questa situazione.

Definizione 2 Una funzione valore è una funzione $v : F \rightarrow \mathbb{R}$ che associa ad ogni impatto un valore reale. Diciamo che v è conforme a una relazione di preferenza Π quando

$$f \preceq f' \Leftrightarrow v(f) \geq v(f') \text{ per ogni } f, f' \in F \quad (3.1)$$

ovvero, equivalentemente

$$\Pi = \{(f, f') \in F \times F : v(f) \geq v(f')\}$$

Le funzioni valore vengono anche definite *funzioni di utilità*: la prima dizione è più matematica, la seconda più economica.

È facile vedere che in questo caso le relazioni derivate di indifferenza e preferenza stretta corrispondono all’uguaglianza e alla differenza stretta fra i valori di v . Il legame con i numeri reali garantito dall’esistenza di una funzione valore conforme induce proprietà forti sulla relazione di preferenza.

Teorema 1 Se una relazione di preferenza ammette una funzione valore conforme, allora la relazione è di ordine debole.

La dimostrazione è semplice, ed è un buon esercizio per verificare la propria comprensione: si tratta di mostrare che la preferenza gode della proprietà riflessiva (perché Π contiene tutte le coppie (f, f) per ogni $f \in F$), della proprietà transitiva (perché per ogni terna $f, f', f'' \in F$ tale che la relazione Π contenga le coppie (f, f') e (f', f'') , essa contiene anche la coppia (f, f'')), e della proprietà di completezza (perché per ogni coppia (f, f') non appartenente a Π , la coppia (f', f) vi appartiene).

Il fatto che una funzione valore induca un ordine debole è interessante, ma non particolarmente utile in pratica, dato che nei problemi decisionali complicati il decisore non ha in mente fin dal principio una funzione valore. Al contrario, si intervista il decisore per capire quali proprietà abbia la sua relazione di preferenza. È quasi certo che sia riflessiva, probabile che sia transitiva. Se il decisore ha le idee chiare, la preferenza potrebbe anche essere completa. In tale caso, sarebbe comodo poterne dedurre che esiste una funzione valore conforme (anche se ancora sconosciuta), perché il problema decisionale si potrebbe ricondurre alla massimizzazione della funzione valore. Siccome questo non è sempre vero, è importante capire quando lo è.

Ordini deboli non riconducibili a una funzione valore

L'ordine lessicografico è un esempio di relazione di ordine debole (anzi, totale) che non ammette una funzione valore conforme. Consideriamo il caso più semplice, cioè quello bidimensionale: gli impatti sono vettori con due componenti reali ($F = \mathbb{R}^2$) e la preferenza è definita come:

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \preceq \begin{bmatrix} f'_1 \\ f'_2 \end{bmatrix} \Leftrightarrow f_1 < f'_1 \text{ oppure } f_1 = f'_1 \text{ e } f_2 \leq f'_2$$

cioè il decisore preferisce l'impatto con la prima componente minore, qualunque sia il valore della seconda; in caso di parità (e solo allora), preferisce l'impatto con la seconda componente minore. Si può dimostrare che questa relazione non ammette nessuna funzione $v(f)$ che rispetti la condizione (3.1), cioè assegna ad ogni impatto in F un valore reale tale che la preferenza fra impatti corrisponda a una relazione fra i valori della funzione.

A livello intuitivo, il motivo è che l'ordine lessicografico richiede la totale incommensurabilità delle due componenti: nessun miglioramento della seconda può compensare un peggioramento anche minimo della prima. Una funzione valore che dipenda da entrambe le componenti, invece, consentirebbe qualche forma di compensazione fra loro. Vedremo che, se la prima componente non fosse continua, ma procedesse per variazioni discrete, diventerebbe possibile costruire una funzione valore conforme che rispetti questa incommensurabilità, dato che non esisterebbero variazioni della prima componente abbastanza piccole da essere compensate da variazioni grandi della seconda.

Ordini deboli riconducibili a una funzione valore³

La maggior parte degli ordini deboli godono di condizioni tecniche aggiuntive che consentono di costruire una funzione valore conforme. Per introdurre tali condizioni occorre una definizione ausiliaria.

³Questa sezione è un approfondimento con argomenti avanzati che non fanno parte del programma.

Definizione 3 *Dato un insieme F e una relazione d'ordine debole Π su F , un sottoinsieme $\bar{F} \subseteq F$ è denso in F quando ogni coppia $f, f' \in F \setminus \bar{F}$ tale che $f \prec f'$ ammette un elemento $\tilde{f} \in \bar{F}$ tale che $f \prec \tilde{f} \prec f'$.*

In altre parole, un sottoinsieme è denso nell'insieme se per ogni coppia di elementi dell'insieme contiene un elemento con preferenza intermedia.

Teorema 2 *Una relazione Π di ordine debole su F ammette una funzione valore conforme in F se e solo se esiste un sottoinsieme $\bar{F} \subseteq F$ numerabile e denso in F .*

L'idea intuitiva è che sia possibile costruire una funzione valore a valori razionali su \bar{F} e poi estenderla su tutto F sfruttando la proprietà di densità, secondo la quale ogni coppia di impatti in F ha un impatto intermedio che appartiene a \bar{F} , e dunque è valorizzabile.

Torniamo all'esempio dell'ordine lessicografico e studiamolo alla luce del concetto di sottoinsieme denso e numerabile. Se $F = \mathbb{R}^2$, nessun sottoinsieme \bar{F} può essere contemporaneamente numerabile e denso in F . Per assurdo, supponiamo che esista un tale sottoinsieme \bar{F} : la proiezione di \bar{F} sull'asse f_1 è un insieme di numeri reali formato da tutte le prime componenti dei vettori di \bar{F} . Siccome \bar{F} è numerabile, ogni suo elemento ha una proiezione e alcune di tali proiezioni possono coincidere, l'insieme delle proiezioni è anch'esso numerabile. Ciò significa che esistono infiniti valori di f_1 che non appartengono all'insieme delle proiezioni. Scegliamo uno di tali valori, \tilde{f}_1 , e scegliamo due valori reali qualsiasi, \tilde{f}_2 e \tilde{f}'_2 . Con questi valori costruiamo i due impatti $[\tilde{f}_1 \tilde{f}_2]^T$ e $[\tilde{f}_1 \tilde{f}'_2]^T$, che appartengono per definizione a $F \setminus \bar{F}$. Se \bar{F} è denso, deve contenere un elemento intermedio fra i due impatti. Per la definizione di ordine lessicografico, tutti gli impatti con preferenza compresa fra questi devono avere una prima componente pari a \tilde{f}_1 e una seconda componente compresa fra \tilde{f}_2 e \tilde{f}'_2 . D'altra parte, per costruzione nessun impatto in \bar{F} può avere prima componente pari a \tilde{f}_1 . La contraddizione porta a concludere che non esiste un sottoinsieme numerabile e denso in F , e quindi non esiste su F una funzione valore conforme.

Cambiando leggermente la definizione di F , questo diventa invece possibile. Supponiamo che l'insieme degli impatti sia $F = \mathbb{N} \times \mathbb{R}$, cioè che la prima componente sia intera. In tal caso, è possibile costruire sottoinsiemi densi e numerabili di F , e quindi trovare funzioni valore conformi all'ordine lessicografico. Un esempio è la funzione $v(f_1, f_2) = f_1 + \tanh(f_2)/2$: il termine f_1 ha valori interi, mentre il termine $\tanh(f_2)/2$ traduce i valori reali di f_2 in valori compresi nell'intervallo aperto $(-1/2, +1/2)$.

Costruire una funzione valore

Non tratteremo completamente il tema di come sia possibile costruire una funzione valore, sotto le condizioni prima introdotte. Ci limiteremo a considerare il caso più semplice in cui F è finito. In tal caso, è sempre possibile definire la funzione denominata *conto di Borda*⁴:

$$v(f) = |\{f' \in F : f \preceq f'\}| \quad f \in F \quad (3.2)$$

Il valore di un impatto è pari al numero di impatti cui esso è preferibile, compreso l'impatto stesso.

Esempio 5 *Si supponga di dover compiere un viaggio e di dover scegliere quale mezzo usare, avendo cinque alternative: $X = \{\text{Aereo, Auto, Pullman, Treno, Taxi}\}$*

⁴Dal nome del matematico francese che lo inventò, Jean-Charles de Borda.

e la relazione di preferenza $\Pi = \{(Aereo, Aereo), (Aereo, Pullman), (Aereo, Taxi), (Auto, Auto), (Auto, Pullman), (Auto, Taxi), (Auto, Treno), (Taxi, Taxi), (Treno, Treno)\}$. Per definizione, la relazione di preferenza dovrebbe contenere coppie di impatti, non di alternative, ma, se assumiamo che ogni alternativa abbia un suo impatto diverso dagli altri, per semplicità possiamo indicare ogni impatto con lo stesso nome dell'alternativa corrispondente.

La relazione Π è rappresentata dal grafo in Figura 3.4. Il conto di Borda permette di costruire una funzione valore conforme alla relazione: l'alternativa *Taxi* è preferibile solo a sé stessa, per cui il valore $v(f(\text{Taxi})) = 1$; le alternative *Treno* e *Pullman* sono fra loro indifferenti e hanno valore 3; l'alternativa *Aereo* ha valore 4 e l'alternativa *Auto*, preferibile a tutte, ha valore 5 (vedi Tabella 15.3).

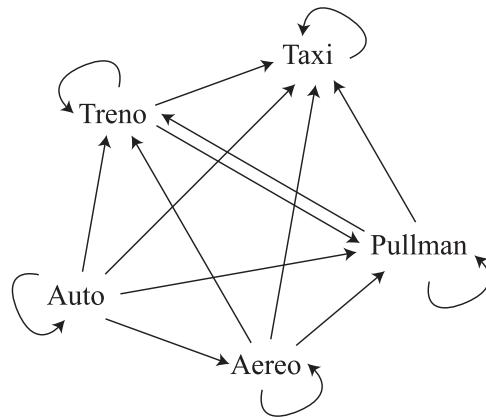


Figura 3.4: Grafo delle preferenze per l'Esempio 5

X	$v(f(x))$
Aereo	4
Auto	5
Pullman	3
Taxi	1
Treno	3

Tabella 3.3: Conto di Borda per l'Esempio 5

Altri casi in cui è facile costruire una funzione valore riguardano impatti monodimensionali ($F \subseteq \mathbb{R}$):

- se f rappresenta un profitto, $u(f) = f$ è una funzione valore conforme;
- se f rappresenta un costo, $u(f) = -f$ è una funzione valore conforme;
- se i valori preferibili di f sono quelli più vicini a un dato valore \bar{f} (ad esempio, f è una temperatura desiderata o un livello di irrigazione ideale per un campo), $u(f) = -(f - \bar{f})^2$ è una funzione valore conforme.

Anche nel caso monodimensionale, però, vi possono essere situazioni complesse.

Esempio 6 Si supponga di dover decidere con quale velocità x correre a prendere un autobus che passa a intervalli regolari. La velocità ammissibile è compresa fra un minimo (il passo più tranquillo che consente di arrivare a destinazione in un tempo ragionevole) e un massimo (che corrisponde alla corsa più rapida considerata accettabile). L'impatto $f(x)$ potrebbe essere l'ora di arrivo alla fermata. Interrogando il decisore, si può scoprire che preferisce arrivare al più presto a destinazione, ma detesta aspettare l'autobus. Di conseguenza, fra due valori di f molto vicini è in generale preferibile il maggiore, perché arrivare più tardi alla fermata riduce l'attesa del bus, ma per alcune coppie critiche di valori di f è nettamente preferibile il minore, dato che consente di prendere il bus precedente, praticamente annullando l'attesa. Quindi, la funzione valore $v(f)$ ha un andamento a dente di sega: cresce gradualmente con f , cala bruscamente, riprende a crescere, cala bruscamente, ecc... Talvolta, un'analisi più approfondita consente di scomporre l'impatto in componenti distinte. Ad esempio, nel caso in questione l'impatto più corretto non è l'ora di arrivo alla fermata, ma una coppia costituita dall'ora di arrivo alla destinazione finale e dalla durata dell'attesa. Però tale impatto non è più espresso da un solo numero⁵.

Molteplicità delle funzioni valore conformi

L'esistenza di un modo per costruire una funzione valore conforme non significa affatto che esista una sola funzione.

Teorema 3 Data una funzione valore $v : F \rightarrow \mathbb{R}$ conforme a una relazione di preferenza Π su F , per qualsiasi funzione strettamente crescente $\phi : \mathbb{R} \rightarrow \mathbb{R}$ la funzione composta $\phi(v(\cdot))$ è anch'essa conforme a Π .

Quindi, se una relazione di ordine debole ammette una funzione valore conforme, ne ammette infinite fra loro equivalenti, che associano valori diversi agli stessi impatti, ma ordinati allo stesso modo. La Tabella 6.1 riporta due funzioni valore conformi alla preferenza descritta nell'Esempio 5: la prima è il conto di Borda, mentre la seconda è del tutto indipendente, ma corrisponde alla stessa relazione di preferenza, allo stesso grafo (quello di Figura 3.4) e allo stesso ordinamento delle alternative.

X	$v(f(x))$	$v'(f(x))$
Aereo	4	20
Auto	5	30
Pullman	3	10
Taxi	1	5
Treno	3	10

Tabella 3.4: Due funzioni valore equivalenti conformi alla relazione di preferenza dell'Esempio 5: v è il conto di Borda

⁵Non sono molto soddisfatto di questo esempio, ma non mi viene in mente nulla di meglio per descrivere un impatto che non sia banalmente un profitto, un costo o la distanza da un valore desiderato: occorre una funzione multimodale con un significato pratico ragionevole.

3.6 Dominanza fra soluzioni

Possiamo ora cominciare a descrivere in che modo i concetti introdotti nelle sezioni precedenti si possano applicare alla scelta di una soluzione per il problema decisionale.

Cominciamo con un caso particolare: supponiamo che il problema abbia un solo decisore e sia deterministico, cioè ammetta un solo scenario ($|D| = |\Omega| = 1$). In tal caso, l'impatto dipende solo dall'alternativa scelta ($f : X \rightarrow F$) e la relazione di preferenza fra impatti induce per proiezione una relazione fra soluzioni: precisamente, una soluzione è preferibile a un'altra quando l'impatto della prima è preferibile a quello della seconda.

$$x \preceq x' \Leftrightarrow f(x) \preceq f(x') \quad x, x' \in X$$

Supponiamo ora che la preferenza Π sia una relazione di preordine, cioè riflessiva e transitiva. La relazione che essa induce sulle soluzioni conserva le stesse proprietà. Osserviamo che anche la completezza si conserva passando da Π alla relazione indotta su X , mentre la proprietà antisimmetrica (se valeva su Π) potrebbe non conservarsi, perché lo stesso impatto può derivare da soluzioni diverse, e quindi due impatti indifferenti sono identici, ma due soluzioni differenti non necessariamente sono identiche.

È ragionevole pensare che, quando un impatto ammette altri impatti che gli sono strettamente preferibili, tutte le soluzioni che lo generano si possano scartare, limitando il problema alle soluzioni i cui impatti non ammettono impatti strettamente preferibili. Nel caso di una preferenza debole, ovviamente, questo non vale, perché non vogliamo che due soluzioni con impatti indifferenti si scartino l'un l'altra (almeno una dovrebbe essere conservata).

Definizione 4 Diremo soluzione dominata una soluzione $x \in X$ tale che $\exists x' \in X : f(x') \prec f(x)$, soluzione non dominata ogni altra soluzione. Infine, diremo $X^* \subseteq X$ l'insieme delle soluzioni non dominate.

Quindi, un decisore razionale sceglie sempre una soluzione non dominata.

Teorema 4 Se l'insieme delle soluzioni non è vuoto ($X \neq \emptyset$) e la relazione di preferenza Π è riflessiva e transitiva, l'insieme delle soluzioni non dominate X^* è non vuoto.

Questo è importante, perché garantisce che il problema ammetta soluzioni ragionevoli. In alcuni casi X^* conterrà una sola soluzione oppure più soluzioni fra loro indifferenti, in altri casi invece conterrà più soluzioni fra loro incomparabili. In generale, quindi, determinare l'insieme X^* semplifica il problema, ma non lo risolve completamente.

Determinazione delle soluzioni non dominate

Se X è finito, si può costruire il grafo delle preferenze strette, i cui nodi corrispondono alle soluzioni, mentre gli archi corrispondono alle coppie di soluzioni i cui impatti sono legati da una preferenza stretta. In questo grafo, non ci sono dunque coppie indifferenti. Le soluzioni non dominate corrispondono ai nodi privi di archi entranti. Per determinarle, basta scorrere gli archi, determinare il grado entrante di ciascun nodo e individuare quelli con grado entrante nullo. Se la complessità di un confronto fra impatti è $O(\gamma)$ (non necessariamente costante, dato che gli impatti $f(x)$ vanno

calcolati, e confrontati considerando tutte le loro componenti), la complessità totale è $O(\gamma |X|^2)$. Ovviamente, questa procedura non è possibile nel caso di problemi infiniti e può essere impraticabile nel caso di problemi combinatorici.

Esempio 7 Riprendiamo l'esempio del viaggio con cinque mezzi alternativi, dove $X = \{Treno, Pullman, Auto, Taxi, Aereo\}$ e $\Pi = \{(Aereo, Aereo), (Aereo, Pullman), (Aereo, Taxi), (Auto, Auto), (Auto, Pullman), (Auto, Taxi), (Auto, Treno), (Taxi, Taxi), (Treno, Taxi), (Treno, Treno)\}$. Il grafo delle preferenze strette è riportato in Figura 3.5: ci sono due soluzioni non dominate, che formano l'insieme $X^* = \{Auto, Aereo\}$.

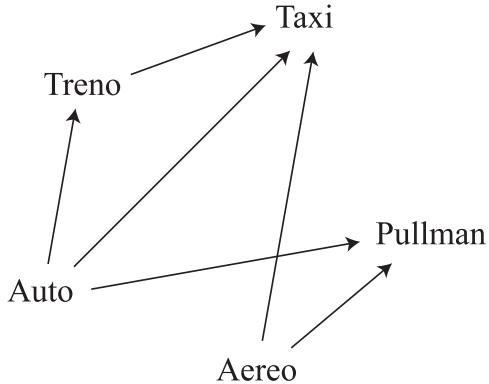


Figura 3.5: Grafo delle preferenze strette per l'Esempio 5

3.7 Esercizi

Esercizio 1

Dato un problema decisionale con insieme di impatti $F = \{a, b, c, d, e, f\}$ e relazione di preferenza $\Pi = \{(a, a), (a, b), (a, c), (a, d), (b, b), (b, c), (b, d), (c, b), (c, c), (c, d), (d, d), (e, b), (e, c), (e, d), (e, e), (f, b), (f, c), (f, d), (f, f)\}$

- si rappresenti la relazione attraverso il grafo corrispondente;
- si elenchino le proprietà di cui Π gode;
- si riconosca se Π è una relazione d'ordine di qualche tipo e che cosa questo consenta di assumere riguardo il problema decisionale;
- si costruiscano le relazioni derivate di indifferenza Ind_{Π} , di preferenza stretta Str_{Π} e di incomparabilità Inc_{Π} .

Soluzione

DA AGGIUNGERE

Esercizio 2

Dato un problema decisionale con insieme di impatti $F = \{a, b, c, d, e\}$ e relazione di preferenza $\Pi = \{(a, a), (a, b), (a, d), (a, e), (b, b), (c, a), (c, b), (c, c), (c, d), (c, e), (d, d), (d, b), (e, a), (e, b), (e, d), (e, e)\}$

- si rappresenti la relazione attraverso il grafo corrispondente;
- si elenchino le proprietà di cui Π gode;
- si riconosca se Π è una relazione d'ordine di qualche tipo e che cosa questo consenta di assumere riguardo il problema decisionale;
- si costruiscano le relazioni derivate di indifferenza Ind_{Π} , di preferenza stretta Str_{Π} e di incomparabilità Inc_{Π} .

Soluzione

DA AGGIUNGERE

Esercizio 3

Dato un problema decisionale con insieme di impatti $F = \{a, b, c, d, e\}$ e relazione di preferenza $\Pi = \{(a, a), (a, b), (a, d), (a, e), (b, b), (c, a), (c, b), (c, c), (c, d), (c, e), (d, d), (d, b), (e, a), (e, b), (e, d), (e, e)\}$

- si rappresenti la relazione attraverso il grafo corrispondente;
- si elenchino le proprietà di cui Π gode;
- si riconosca se Π è una relazione d'ordine di qualche tipo e che cosa questo consenta di assumere riguardo il problema decisionale;
- si costruiscano le relazioni derivate di indifferenza Ind_{Π} , di preferenza stretta Str_{Π} e di incomparabilità Inc_{Π} .

Soluzione

DA AGGIUNGERE

Esercizio 4

Dato un problema decisionale con insieme di impatti $F = \{a, b, c, d, e\}$ e relazione di preferenza $\Pi = \{(a, a), (a, b), (a, c), (a, d), (a, e), (b, b), (b, c), (b, d), (b, e), (c, b), (c, c), (c, d), (c, e), (d, d), (e, e)\}$

- si rappresenti la relazione attraverso il grafo corrispondente;
- si elenchino le proprietà di cui Π gode;
- si riconosca se Π è una relazione d'ordine di qualche tipo e che cosa questo consenta di assumere riguardo il problema decisionale;
- si costruiscano le relazioni derivate di indifferenza Ind_{Π} , di preferenza stretta Str_{Π} e di incomparabilità Inc_{Π} .

Soluzione

DA AGGIUNGERE

Esercizio 5

Dato un problema decisionale con la relazione di preferenza Π descritta dalla seguente tabella:

Π	f	f'	f''	f'''	f''''
f	1	0	1	1	1
f'	1	1	0	1	1
f''	0	0	1	1	1
f'''	0	0	0	1	0
f''''	0	0	1	0	1

- si rappresenti la relazione attraverso il grafo corrispondente;
- si elenchino le proprietà di cui Π gode;
- si riconosca se Π è una relazione d'ordine di qualche tipo e che cosa questo consenta di assumere riguardo il problema decisionale;
- si costruiscano le relazioni derivate di indifferenza Ind_{Π} , di preferenza stretta Str_{Π} e di incomparabilità Inc_{Π} .

Soluzione

DA AGGIUNGERE

Capitolo 4

Programmazione Matematica

Questo capitolo tratta la categoria più semplice dei problemi decisionali:

- il sistema è deterministico, cioè c'è un solo scenario:

$$\Omega = \{\bar{\omega}\}$$

e, siccome le variabili esogene sono costanti ($\omega_k = \bar{\omega}_k$), l'impatto f dipende solo dall'alternativa x e si può indicare semplicemente con $f(x)$;

- c'è un solo decisore:

$$|D| = 1$$

- la relazione di preferenza Π ammette una funzione valore conforme nota:

$$\exists v : F \rightarrow \mathbb{R} : \Pi = \{(f, f') \in F \times F : v(f) \geq v(f')\}$$

Questo tipo di problemi viene trattato soprattutto nella letteratura matematica, dove prevale l'uso di esprimere la preferenza con una funzione di costo, anziché con una funzione valore. La differenza è che tale funzione associa agli impatti preferiti valori inferiori anziché superiori, e va quindi minimizzata, anziché massimizzata. Per brevità, indicheremo direttamente con f il costo anziché l'impatto della soluzione, scrivendo:

$$\begin{aligned} & \min f(x) \\ & x \in X \end{aligned}$$

Faremo anche un'ipotesi aggiuntiva sulla struttura dell'insieme delle soluzioni X , cioè che esso sia descrivibile attraverso un insieme finito di diseguaglianze:

$$X = \{x \in \mathbb{R}^n : g_j(x) \leq 0 \text{ per } j = 1, \dots, m\}$$

in cui $g_j : X \rightarrow \mathbb{R}$ e $g_j \in C^1(X)$ per $j = 1, \dots, m$, cioè le funzioni g_j sono funzioni a valori reali continue con la loro derivata prima nell'insieme delle soluzioni ammissibili del problema. Infine, assumeremo che anche la funzione di costo sia continua con la sua derivata prima nello stesso insieme: $f \in C^1(X)$.

Questo insieme di problemi (a volte con ipotesi più o meno strette sulla continuità delle funzioni e delle derivate) vanno sotto l'ampio nome di problemi di *Programmazione Matematica*. Il motivo è che descrivere sia la preferenza sia la regione ammissibile attraverso funzioni a valori reali consente di applicare a questi problemi decisionali tutto l'armamentario dell'analisi matematica.

4.1 Concetti di base

Definizione 5 Dato un insieme $X \subseteq \mathbb{R}^n$ e una funzione $f : X \rightarrow \mathbb{R}$, definiamo punto di ottimo globale un punto $x^\circ \in X$ tale che

$$f(x^\circ) \leq f(x) \quad \text{per ogni } x \in X$$

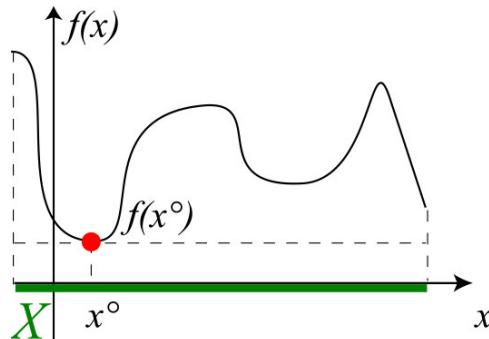


Figura 4.1: Grafico di una funzione $f(x)$ con la regione ammissibile X e un punto di ottimo globale x°

La Figura 4.1 mostra un esempio di funzione definita in un intervallo e dotata di un punto di ottimo globale. Dalle considerazioni della Sezione 3.6, appare chiaramente che le soluzioni di un problema decisionale che ricade nel campo della programmazione matematica sono i punti di ottimo globale. Siccome determinare tali punti non è facile, consideriamo un secondo insieme di punti, più ampio, per il quale è invece possibile sviluppare metodi di calcolo.

Definizione 6 Dato un insieme $X \subseteq \mathbb{R}^n$ e una funzione $f : X \rightarrow \mathbb{R}$ definiamo punto di ottimo locale un punto $x^* \in X$ tale che

$$\exists \epsilon > 0 : f(x^*) \leq f(x) \quad \text{per ogni } x \in X \cap \mathcal{U}_{x^*, \epsilon}$$

dove $\mathcal{U}_{x^*, \epsilon} = \{x \in \mathbb{R}^n : \|x - x^*\| < \epsilon\}$ è un intorno di x^* di ampiezza ϵ .

Osservazione 1 Un punto di ottimo globale è per definizione anche un punto di ottimo locale:

$$X^\circ \subseteq X^*$$

La Figura 4.1 mostra un esempio di funzione definita in un intervallo e dotata di vari punti di ottimo locale, uno dei quali è anche punto di ottimo globale.

Siccome non esistono metodi generali per trovare X^* né X° , cercheremo invece delle condizioni necessarie per l'ottimalità locale. Queste condizioni individuano punti candidati a essere punti di ottimo locale, e quindi di ottimo globale. In questo modo, indeboliamo la richiesta due volte:

$$\begin{array}{ccccccc} \text{Ottimo globale} & \Rightarrow & \text{Ottimo locale} & \Rightarrow & \text{Condizioni di KKT} \\ X^\circ & \subseteq & X^* & \subseteq & X^{\text{KKT}} \end{array}$$

Le condizioni che tratteremo sono dette *condizioni di Karush-Kuhn-Tucker* (in breve, *condizioni di KKT*), dai nomi dei loro scopritori¹.

¹William Karush (1917-1997) scoprì le condizioni nel 1939 nella sua tesi di laurea, senza pubblicarle; Harold William Kuhn (1925-2014) e Albert William Tucker (1905-1995) le riscoprirono nel 1951, quando nessuno ancora conosceva il lavoro di Karush.

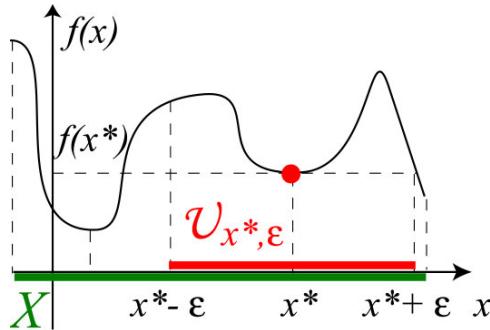


Figura 4.2: Grafico di una funzione $f(x)$ con la regione ammissibile X e un punto di ottimo locale x^* (gli altri due sono x° e il massimo dell'insieme X)

Procederemo in questo modo:

1. imposteremo le condizioni come sistema di uguaglianze e disegualanze;
2. risolveremo il sistema, per ottenere l'insieme X^{KKT} dei punti candidati;
3. valuteremo uno per uno i punti di X^{KKT} per conservare solo i migliori.

I punti migliori di X^{KKT} formeranno X° . Il metodo richiede che X^{KKT} sia un insieme finito o almeno se è descrivibile analiticamente, in modo da poter determinare X° .

Le condizioni di KKT si basano sull'approssimare sia la funzione obiettivo sia le funzioni che descrivono i vincoli con funzioni lineari delle variabili di decisione, e dedurne delle condizioni necessarie che tali funzioni approssimanti devono rispettare in un punto \tilde{x} dato affinché \tilde{x} sia un punto di ottimo locale.

4.1.1 Sviluppo in serie di Taylor

Ogni funzione sufficientemente regolare (cioè continua con tutte le sue derivate fino a un ordine k) si può approssimare localmente con un polinomio di grado k .

Teorema 5 *Sia $f \in C^k(\mathcal{U}_{\tilde{x},\epsilon})$ una funzione di una variabile $x \in \mathbb{R}$. Per ogni $x \in \mathcal{U}_{\tilde{x},\epsilon}$, vale lo sviluppo in serie di Taylor:*

$$f(x) = \sum_{i=0}^k f^{(i)}(\tilde{x})(x - \tilde{x})^i + R_k(x - \tilde{x})$$

$$\text{con } f^{(0)} = f, \quad f^{(i)} = \frac{d^i f}{dx^i} \text{ per ogni } i > 0, \quad \text{e } \lim_{x \rightarrow \tilde{x}} \frac{R_k(x - \tilde{x})}{\|x - \tilde{x}\|^k} = 0.$$

La condizione sul limite di R_k significa che l'errore di approssimazione commesso considerando solo i primi $k+1$ termini, cioè il polinomio di grado k , è piccolo per $x \approx \tilde{x}$, e soprattutto tende a zero più velocemente di quanto x tenda a \tilde{x} . Questa proprietà permetterà di ignorare tale errore in molti procedimenti di passaggio al limite. Noi useremo solo lo sviluppo fino a $k=1$; qualsiasi funzione regolare fino al primo ordine ammette un'approssimazione lineare:

$$f(x) = f(\tilde{x}) + f'(\tilde{x})(x - \tilde{x}) + R_1(|x - \tilde{x}|)$$

che si può generalizzare da funzioni di una sola variabile a funzioni di più variabili ($x \in \mathbb{R}^n$) scrivendo²:

$$f(x) = f(\tilde{x}) + (\nabla f(\tilde{x}))^T (x - \tilde{x}) + R_1(\|x - \tilde{x}\|) \quad (4.1)$$

4.1.2 Direzioni

Dato un punto $\tilde{x} \in \mathbb{R}^n$, un vettore $d \in \mathbb{R}^n$ si dice *direzione* se è pensato come un termine da aggiungere a \tilde{x} per spostarsi in altri punti, modulandone la lunghezza con un opportuno coefficiente α in modo da decidere di quanto spostarsi.

Definizione 7 Una direzione ammissibile in $\tilde{x} \in X$ è un vettore $d \in \mathbb{R}^n$ tale che

$$\exists \bar{\alpha} > 0 : \tilde{x} + \alpha d \in X \quad \forall \alpha \in [0; \bar{\alpha})$$

In altre parole, spostandosi da \tilde{x} abbastanza poco (meno di $\bar{\alpha}$) in direzione d , la soluzione raggiunta è ancora ammissibile.

Definizione 8 Una direzione migliorante in $\tilde{x} \in X$ è un vettore $d \in \mathbb{R}^n$ tale che

$$\exists \bar{\alpha} > 0 : f(\tilde{x} + \alpha d) < f(\tilde{x}) \quad \forall \alpha \in (0; \bar{\alpha})$$

In altre parole, spostandosi da \tilde{x} abbastanza poco (meno di $\bar{\alpha}$) in direzione d , la soluzione raggiunta è migliore di quella iniziale.

4.2 Condizioni necessarie di ottimalità locale

Questa sezione introduce una condizione necessaria per l'ottimalità locale, ne deriva un'intuizione algoritmica su come trovare punti candidati e riformula la condizione in modo da rendere l'intuizione algoritmica praticamente realizzabile.

Teorema 6 Se \tilde{x} è un punto di ottimo locale e d è una direzione ammissibile in \tilde{x} , d non è una direzione migliorante in \tilde{x} .

Dimostrazione. Se \tilde{x} è un punto di ottimo locale, allora $f(x) \geq f(\tilde{x})$ per ogni $x \in \mathcal{U}_{\tilde{x}, \epsilon} \cap X$. Per assurdo, supponiamo che esista una direzione d insieme ammissibile e migliorante. Allora, esistono $\bar{\alpha}_1$ e $\bar{\alpha}_2 > 0$ tali che:

- $\tilde{x} + \alpha d \in X$ per ogni $\alpha \in [0; \bar{\alpha}_1)$
- $f(\tilde{x} + \alpha d) < f(\tilde{x})$ per ogni $\alpha \in (0; \bar{\alpha}_2)$

Ora poniamo $\alpha = \frac{1}{2} \min(\epsilon / \|d\|, \bar{\alpha}_1, \bar{\alpha}_2) > 0$ e $x = \tilde{x} + \alpha d$. Dai tre termini dell'operatore di minimo deriva che:

- $\|x - \tilde{x}\| = \|\alpha d\| = \alpha \|d\| \leq \frac{1}{2} \epsilon < \epsilon \Rightarrow x \in \mathcal{U}_{\tilde{x}, \epsilon}$
- $x \in X$
- $f(x) < f(\tilde{x})$

ma poiché $x \in X \cap \mathcal{U}_{\tilde{x}, \epsilon}$ deve essere anche $f(x) \geq f(\tilde{x})$, che è un assurdo. ■

²Per le definizioni dei simboli, si rimanda all'Appendice A, relativa ai richiami di analisi matematica.

Un approccio algoritmico

Il Teorema 6 suggerisce un approccio algoritmico:

- considerare tutti i punti ammissibili come candidati;
- scorrere tutti i punti candidati;
- per ogni punto ammissibile \tilde{x} , scorrere tutte le direzioni ammissibili;
- per ogni direzione d ammissibile in \tilde{x} , valutare se la direzione è migliorante: se lo è, cancellare il punto dall'insieme dei candidati;
- al termine, scorrere i punti candidati, conservando solo i migliori.

Questo è un finto algoritmo, dato che l'insieme dei punti ammissibili e quello delle direzioni ammissibili in generale sono infiniti; inoltre, per valutare se una direzione è migliorante, bisogna analizzare la funzione $f(\tilde{x} + \alpha d)$ su un intero intervallo di valori di α . È un metodo persino peggiore di quello esaustivo che scorre tutti i punti ammissibili e conserva i migliori (anch'esso un procedimento infinito). Per rendere praticabile questo procedimento, bisogna sostituire le varie condizioni che lo compongono con condizioni calcolabili in un tempo finito. Queste condizioni dovrebbero essere equivalenti, ma siccome ciò non è sempre possibile, bisognerà fare una serie di precisazioni. Le condizioni calcolabili verranno ottenute sfruttando l'approssimazione lineare della funzione obiettivo f e dei vincoli g_j fornita dallo sviluppo in serie di Taylor.

Nelle prossime sezioni procederemo come segue:

1. sostituiremo la condizione di non miglioramento dell'obiettivo con una condizione sul gradiente della funzione obiettivo;
2. sostituiremo la condizione di ammissibilità con una condizione sui gradienti dei vincoli (a questo punto, daremo una prima interpretazione geometrica delle condizioni rimaneggiate);
3. ribalteremo queste condizioni, applicando un opportuno lemma per ottenere le condizioni di KKT, che legano il gradiente dell'obiettivo e quelli dei vincoli (anche di queste condizioni daremo un'interpretazione geometrica).

4.2.1 Condizione di non miglioramento

Il primo passaggio riguarda la funzione obiettivo: siccome è difficile verificare che una data direzione d sia non migliorante in \tilde{x} , scriviamo una condizione molto più semplice da verificare. Tale condizione non è equivalente, ma solo necessaria, dato che è una conseguenza della prima. Quindi, i punti in cui essa vale non sono punti di ottimo locale, ma semplicemente punti candidati.

Teorema 7 *Sia $f \in C^1(X)$, $\tilde{x} \in X$ una soluzione ammissibile e d una direzione ammissibile e non migliorante in \tilde{x} . Allora:*

$$(\nabla f(\tilde{x}))^T d \geq 0$$

Dimostrazione. Se d è direzione ammissibile e non migliorante in \tilde{x} , esiste un valore $\bar{\alpha}$ tale che $f(\tilde{x} + \alpha d) \geq f(\tilde{x})$ per ogni $\alpha \in [0; \bar{\alpha}]$. Per lo sviluppo in serie di Taylor è:

$$\begin{aligned} f(\tilde{x}) + (\nabla f(\tilde{x}))^T (\tilde{x} + \alpha d - \tilde{x}) + R_1(\|\alpha d\|) &\geq f(\tilde{x}) \Rightarrow \\ \Rightarrow (\nabla f(\tilde{x}))^T d + \frac{R_1(\|\alpha d\|)}{\alpha} &\geq 0 \end{aligned}$$

In queste disuguaglianze, \tilde{x} , d e il gradiente $\nabla f(\tilde{x})$ sono costanti; l'unica variabile è il parametro α . Ora facciamo tendere α a 0; per continuità, la disuguagliaanza si conserva:

$$\lim_{\alpha \rightarrow 0} \left[(\nabla f(\tilde{x}))^T d + \frac{R_1(\alpha d)}{\alpha \|d\|} \|d\| \right] \geq 0 \Rightarrow (\nabla f(\tilde{x}))^T d \geq 0$$

■ La condizione $(\nabla f(\tilde{x}))^T d \geq 0$ ha un semplice significato geometrico: la direzione d forma un angolo $\leq 90^\circ$ con il gradiente della funzione obiettivo, che è la direzione di più rapido peggioramento (cioè aumento). Vale a dire che d è non migliorante.

Si noti che, per verificare questa condizione basta calcolare $\nabla f(\tilde{x})$, che è un vettore costante, e farne il prodotto scalare con d , che è un dato. Al contrario, applicare la definizione per verificare che d è non migliorante è in pratica impossibile, dato che comporta di verificare una disuguaglianza per infiniti valori di α . D'altra parte, la condizione sul gradiente è vera anche per direzioni che in realtà sono miglioranti. Questo succede quando $(\nabla f(\tilde{x}))^T d = 0$ e il miglioramento è dovuto al resto R_1 . Questo significa che in alcuni casi limite risultano candidati punti che non sono punti di ottimo locale.

Esempio 8 Consideriamo il seguente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 1)^2 + x_2^2 \\ g_1(x) &= -x_1^2 - x_2^2 + 4 \leq 0 \\ g_2(x) &= x_1 - 3/2 \leq 0 \end{aligned}$$

la cui regione ammissibile è riportata in Figura 4.3.

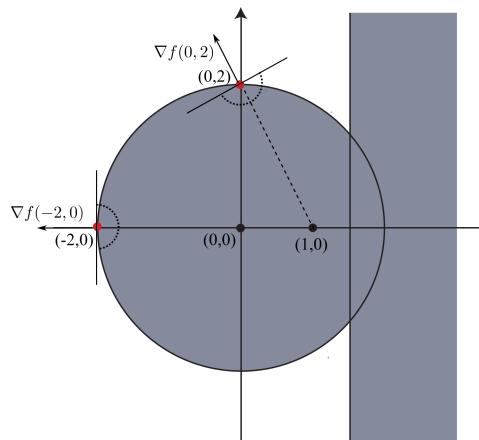


Figura 4.3: Regione ammissibile del problema riportato nell'Esempio 8

Il gradiente della funzione obiettivo è $\nabla f(x) = [2(x_1 - 1) \ 2x_2]^T$ e in ogni punto è diretto ad allontanarsi dal punto $(1, 0)$. Per esempio³, nel punto $\tilde{x} = (0, 2)$ il gradiente è $\nabla f(0, 2) = [-2 \ 4]^T$, mentre nel punto $\tilde{x} = (-2, 0)$ il gradiente è $\nabla f(-2, 0) = [-6 \ 0]^T$. Condizione necessaria perché una direzione sia migliorante è che sia confinata nel semispazio indicato dalla semicirconferenza a puntini. Se ci

³Per semplicità, la figura li rappresenta i gradienti con la stessa lunghezza, perché ai fini della ricerca dei punti candidati conta solo la direzione in cui puntano.

si muove da $\tilde{x} = (0, 2)$ verso destra nella direzione $p = [1 \ 0]^T$, la funzione obiettivo migliora. Infatti

$$(\nabla f(0, 2))^T p = [-2 \ 4] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -2 < 0$$

Invece, muoversi verso l'alto nella direzione $p = [0 \ 1]^T$, la funzione obiettivo non migliora. Infatti

$$(\nabla f(0, 2))^T p = [-2 \ 4] \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4 \geq 0$$

Un caso limite si può osservare nel punto $\tilde{x} = (-2, 0)$: se ci si muove verso il basso con vettore tangente $p = [0 \ -1]^T$ la condizione diventa:

$$(\nabla f(-2, 0))^T p = [-6 \ 0] \begin{bmatrix} 0 \\ -1 \end{bmatrix} = 0$$

In effetti, questa direzione è peggiorante, ma esistono funzioni obiettivo con lo stesso gradiente per le quali essa diventerebbe migliorante (per esempio, $f(x) = -6x_1 + x_2^2$). Per questo, le direzioni limite non possono essere usate per scartare punti candidati.

L'“algoritmo” che ne risulta (fra virgolette perché contiene ancora due cicli infiniti annidati) è riportato in Figura 4.4.

```

Algorithm TrovaCandidati( $f, X$ )
 $X^{\text{KKT}} := X$ ;
For each  $\tilde{x} \in X^{\text{KKT}}$  do
    { Scarta i punti che violano le condizioni necessarie }
    For each  $d \in \mathbb{R}^n$  ammissibile in  $\tilde{x}$  do
        If  $(\nabla f(\tilde{x}))^T d < 0$  then  $X^{\text{KKT}} := X^{\text{KKT}} \setminus \{\tilde{x}\}$ 
    { Restituisce il miglior punto candidato }
    Return  $X^{\text{KKT}}$ ;

```

Figura 4.4: Pseudocodice del procedimento per ridurre l'insieme ammissibile X al sottoinsieme dei punti candidati X^{KKT} .

4.2.2 Condizione di ammissibilità

Molto più problematico è sostituire la condizione che d sia una direzione ammissibile. Ci sono due difficoltà:

1. esistono problemi in cui nessuna direzione è ammissibile, e quindi la condizione per scartare punti candidati non entra mai in gioco;
2. la condizione di ammissibilità è un'ipotesi del Teorema 6, e quindi non si può sostituire con condizioni più deboli, ma solo con condizioni equivalenti o più forti (la condizione di non miglioramento invece è la tesi, e si può tranquillamente indebolire).

Un caso importante in cui si verifica la prima difficoltà è quello dei problemi che contengono vincoli di uguaglianza non lineari.

Esempio 9 Supponiamo che la regione ammissibile sia la circonferenza unitaria centrata nell'origine:

$$X = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$$

Nessuno di questi punti ha direzioni ammissibili. Per esempio, sia $\tilde{x} = (1, 0) \in X$ e supponiamo per assurdo che d sia una direzione ammissibile:

$$\tilde{x} + \alpha d = [1 + \alpha d_1 \quad \alpha d_2]^T \in X \text{ per ogni } \alpha \in [0; \bar{\alpha})$$

Questo richiede $(1 + \alpha d_1)^2 + (\alpha d_2)^2 = 1 \Rightarrow 2\alpha d_1 + \alpha^2 (d_1^2 + d_2^2) = 0$, da cui $\alpha = -d_1 / (d_1^2 + d_2^2)$, cioè α è fissato. Ma la condizione dovrebbe valere per ogni α in un opportuno intervallo. Siccome questo non è possibile, nessuna direzione d è ammissibile e \tilde{x} non si può scartare. Lo stesso vale per ogni altro punto di X : il procedimento termina quindi con $X^{\text{KKT}} = X$.

Archi ammissibili

Fortunatamente, è possibile estendere il Teorema 6 considerando non solo spostamenti rettilinei da \tilde{x} in una direzione fissa, ma anche spostamenti curvilinei (per esempio, lungo un arco di circonferenza). Il risultato di questa generalizzazione è molto simile a quello del Teorema 7.

Definizione 9 Si dice arco ammissibile in un punto $\tilde{x} \in X$ una linea parametrica $\xi : \mathbb{R}^+ \rightarrow \mathbb{R}^n$ che parte da \tilde{x} e appartiene a X :

$$\begin{aligned} \xi(0) &= \tilde{x} \\ \xi(\alpha) &\in X \quad \text{per ogni } \alpha \in [0; \bar{\alpha}) \end{aligned}$$

Definizione 10 Si dice direzione tangente all'arco $\xi(\alpha)$ ammissibile in \tilde{x} il vettore costituito dalle derivate prime delle componenti ξ_i rispetto al parametro α , valutate nel punto iniziale dell'arco:

$$p_\xi = \left[\frac{d\xi_1}{d\alpha} \Big|_0 \cdots \frac{d\xi_n}{d\alpha} \Big|_0 \right]^T$$

Esempio 10 Dato $X = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$, nel punto $\tilde{x} = (1, 0)$ è ammissibile l'arco

$$\xi(\alpha) = \begin{bmatrix} \cos \alpha \\ \sin \alpha \end{bmatrix}$$

dato che $\xi(0) = \tilde{x}$ e $\xi(\alpha) \in X$ per ogni $\alpha \in [0; +\infty)$.

Tale arco ha direzione tangente

$$p_\xi = \begin{bmatrix} -\sin 0 \\ \cos 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Definizione 11 Si dice arco migliorante in un punto $\tilde{x} \in X$ una linea parametrica $\xi : \mathbb{R}^+ \rightarrow \mathbb{R}^n$ che parte da \tilde{x} e per la quale esista un $\bar{\alpha} > 0$ tale che

$$f(\xi(\alpha)) < f(\tilde{x}) \quad \text{per ogni } \alpha \in [0; \bar{\alpha})$$

Direzioni e archi

Gli archi non sono altro che una generalizzazione delle direzioni. Infatti, a ogni direzione d in \tilde{x} si può associare un arco, che è semplicemente la semiretta $\xi(\alpha) =$

$\tilde{x} + \alpha d$. La tangente a tale arco è la direzione stessa: $p_\xi = d$. Quindi, le direzioni (in particolare, quelle ammissibili, miglioranti, ecc...) sono un sottoinsieme (molto piccolo) degli archi o, più precisamente, un sottoinsieme dei vettori tangenti a tali archi.

Il teorema seguente dimostra che si possono filtrare i punti candidati anche con le tangenti ad archi ammissibili, esattamente come con le direzioni ammissibili.

Teorema 8 *Sia $f \in C^1(X)$, $\tilde{x} \in X$ un punto ammissibile e $\xi : \mathbb{R} \rightarrow \mathbb{R}^n$ un arco ammissibile e non migliorante in \tilde{x} . Allora*

$$(\nabla f(\tilde{x}))^T p_\xi \geq 0$$

Dimostrazione. Se $\xi(\alpha)$ è un arco ammissibile, $\exists \bar{\alpha}$ tale che $\xi(\alpha) \in X$ per ogni $\alpha \in [0; \bar{\alpha})$ e

$$\begin{aligned} f(\xi(\alpha)) &\geq f(\tilde{x}) \Rightarrow \\ \Rightarrow f(\xi(0)) + \frac{df}{d\alpha} \Big|_{\alpha=0} \alpha + R_1(\|\xi(\alpha) - \xi(0)\|) &\geq f(\tilde{x}) \Rightarrow \\ \Rightarrow (\nabla f(\tilde{x}))^T p_\xi + \frac{R_1(\|\xi(\alpha) - \xi(0)\|)}{\alpha} &\geq 0 \end{aligned}$$

dove si sono usate le proprietà delle derivate di funzioni composte a più variabili (vedi Appendice A).

Se facciamo tendere α a 0, per continuità la disegualanza si conserva:

$$\lim_{\alpha \rightarrow 0} \left((\nabla f(\tilde{x}))^T p_\xi + \frac{R_1(\xi(\alpha) - \xi(0)) \|\xi(\alpha) - \xi(0)\|}{\alpha} \right) \geq 0 \Rightarrow (\nabla f(\tilde{x}))^T p_\xi \geq 0$$

■

Esempio 11 Riprendiamo l'Esempio 9 sui problemi con vincoli non lineari. Sia $X = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ la regione ammissibile e $f(x) = -x_2$ la funzione obiettivo, con gradiente pari a $(\nabla f(\tilde{x})) = [0 \ -1]^T$ ovunque. Se si considera il punto $\tilde{x} = (1, 0)$ e l'arco ammissibile $\xi(\alpha) = [\cos \alpha \ \sin \alpha]^T$, con vettore tangente $p_\xi = [0 \ 1]^T$, si può verificare che:

$$(\nabla f(\tilde{x}))^T p_\xi = [0 \ -1]^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -1 < 0$$

e quindi si può scartare il punto stesso. Invece nel punto $x^\circ = (0, 1)$ si hanno due archi ammissibili: $\xi(\alpha) = [\sin \alpha \ \cos \alpha]^T$, con vettore tangente $p_\xi = [1 \ 0]^T$ e $\xi'(\alpha) = [-\sin \alpha \ \cos \alpha]^T$, con vettore tangente $p_{\xi'} = [-1 \ 0]^T$ e si può verificare che:

$$(\nabla f(\tilde{x}))^T p_\xi = [0 \ -1]^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0 \quad (\nabla f(\tilde{x}))^T p_{\xi'} = [0 \ -1]^T \begin{bmatrix} -1 \\ 0 \end{bmatrix} = 0$$

Questo è l'unico punto candidato di X ed è il punto di ottimo globale.

Esempio 12 Riprendiamo l'esempio di Figura 4.3. Oltre alle direzioni rettilinee si possono considerare anche archi curvilinei. Il caso più interessante è quello del punto $(-2, 0)$, dove, oltre alla direzione $[0 \ -1]^T$ si può considerare l'arco che scorre lungo la circonferenza di raggio 2 segnata in grigio: questo arco è migliorante, dato che la distanza dal punto $(1, 0)$ diminuisce strettamente, ma la direzione tangente è la stessa, e quindi la condizione di non miglioramento viene soddisfatta esattamente. La conclusione è che, quando le condizioni non sono certe, non si possono usare per scartare punti candidati.

Caratterizzazione degli archi ammissibili

Dover considerare tutti gli archi ammissibili è persino peggio che considerare le direzioni ammissibili. Cerchiamo quindi di esprimere analiticamente la condizione che una data direzione p sia tangente ad un arco ammissibile, esattamente come prima avevamo espresso analiticamente la condizione che una direzione fosse non migliorante. Questo ci permetterà di non scorrere tutti gli archi, valutando quali siano ammissibili e quali no, ma di valutare una condizione più semplice. Ricordiamo che la regione ammissibile è definita come

$$X = \{x \in \mathbb{R}^n : g_j(x) \leq 0, \text{ per } j = 1, \dots, m\}$$

con $g_j \in C^1(X)$ per $j = 1, \dots, m$.

Definizione 12 Si dice vincolo attivo in un punto $\tilde{x} \in X$ qualsiasi vincolo $g_j(x) \leq 0$ tale che $g_j(\tilde{x}) = 0$.

Esempio 13 Consideriamo l'esempio di Figura 4.3. Nel punto $(-2, 2)$ nessun vincolo è attivo; infatti, $g_1(-2, 2) = -4$ e $g_2(-2, 2) = -7/2$. Nel punto $(0, 2)$ è attivo solo il primo vincolo, dato che $g_1(0, 2) = 0$ e $g_2(0, 2) = -3/2$. Nel punto $(3/2, 2)$ è attivo solo il secondo vincolo, dato che $g_1(3/2, 2) = -9/4$ e $g_2(3/2, 2) = 0$. Nel punto $(3/2, \sqrt{7}/2)$ sono attivi entrambi i vincoli, dato che $g_1(3/2, \sqrt{7}/2) = 0$ e $g_2(3/2, \sqrt{7}/2) = 0$.

Teorema 9 Sia $\xi(\alpha)$ un arco ammissibile in $\tilde{x} \in X$ con vettore tangente p_ξ e $g_j \leq 0$ un vincolo attivo in \tilde{x} . Allora

$$(\nabla g_j(\tilde{x}))^T p_\xi \leq 0$$

Dimostrazione. Se $\xi(\alpha)$ è un arco ammissibile in \tilde{x} , esiste $\bar{\alpha} > 0$ tale che

$$g_j(\xi(\alpha)) \leq 0 \quad \text{per ogni } \alpha \in [0; \bar{\alpha}] \text{ e } j = 1, \dots, m$$

ovvero

$$\begin{aligned} g_j(\xi(\alpha)) &= g_j(\xi(0)) + \frac{dg_j}{d\alpha} \Big|_0 \alpha + R_1(\|\xi(\alpha) - \xi(0)\|) = \\ &= g_j(\tilde{x}) + \alpha (\nabla g_j(\tilde{x}))^T p_\xi + R_1(\|\xi(\alpha) - \xi(0)\|) \leq 0 \end{aligned}$$

Per i vincoli non attivi in \tilde{x} , la disegualanza è ovvia perché $g_j(\tilde{x}) < 0$ e gli altri due termini hanno limite nullo per $\alpha \rightarrow 0$. Dunque, per continuità, quando α è sufficientemente piccolo non riescono a ribaltare il segno della disegualanza. Per i vincoli attivi in \tilde{x} , invece, la disegualanza dice qualcosa di molto utile. Infatti, per tali vincoli è $g_j(\tilde{x}) = 0$, per cui

$$g_j(\xi(\alpha)) = \alpha (\nabla g_j(\tilde{x}))^T p_\xi + R_1(\|\xi(\alpha) - \xi(0)\|) \leq 0$$

Dividendo ambo i membri per α e passando al limite, per continuità si ottiene la tesi:

$$\lim_{\alpha \rightarrow 0} \left[(\nabla g_j(\tilde{x}))^T p_\xi + \frac{R_1(\|\xi(\alpha) - \xi(0)\|)}{\|\xi(\alpha) - \xi(0)\|} \frac{\|\xi(\alpha) - \xi(0)\|}{\alpha} \right] = (\nabla g_j(\tilde{x}))^T p_\xi \leq 0$$

■

Questa proprietà è interessante, ma ora subentra il secondo problema elencato nella Sezione 4.2.2: non possiamo sostituirla alla condizione che p_ξ sia direzione tangente ad un arco ammissibile. Se fosse condizione sufficiente o equivalente, potremmo usarla come ipotesi nel Teorema 8 e ridimostrarlo fino ad ottenere la tesi, ma in generale è solo una condizione necessaria. Per fortuna, si può dimostrare che, in situazioni opportune, la condizione è davvero equivalente.

Punti regolari

Definizione 13 Si dice punto regolare per un dato sistema di vincoli $g_j(x) \leq 0$ un punto in cui i vincoli attivi hanno gradienti linearmente indipendenti tra loro (condizione di constraint qualification).

In un punto regolare, le condizioni sui gradienti dei vincoli attivi sono non solo necessarie, ma anche sufficienti a garantire che esista un arco ammissibile con un dato vettore tangente.

Teorema 10 Sia \tilde{x} un punto regolare. Esiste un arco $\xi(\alpha)$ ammissibile in \tilde{x} con direzione tangente $p_\xi = p$ se e solo se

$$(\nabla g_j(\tilde{x}))^T p \leq 0 \quad \text{per ogni } j : g_j(\tilde{x}) = 0$$

Ne deriva che nei punti regolari possiamo riscrivere il Teorema 7 sostituendo le condizioni sui gradienti dei vincoli attivi all'ipotesi di avere un arco ammissibile, cioè possiamo usare le condizioni analitiche sui gradienti per identificare le direzioni utili a filtrare i punti candidati. I punti non regolari vanno trattati a parte. Fortunatamente, in genere sono un'esigua minoranza dei punti ammissibili, e quindi li includeremo d'ufficio nell'insieme dei punti candidati senza ulteriori analisi.

Corollario 1 Sia $f \in C^1(X)$, $\tilde{x} \in X$ un punto regolare di ottimo locale e p un vettore tale che

$$(\nabla f(\tilde{x}))^T p \leq 0 \quad \text{per ogni } j : g_j(\tilde{x}) = 0$$

Allora:

$$(\nabla f(\tilde{x}))^T p \geq 0$$

Esempio 14 Si consideri il problema:

$$\begin{aligned} \min f(x) &= x_2 \\ g_1(x) &= (x_1 - 1)^3 + (x_2 - 2) \leq 0 \\ g_2(x) &= (x_1 - 1)^3 - (x_2 - 2) \leq 0 \\ g_3(x) &= -x_1 \leq 0 \end{aligned}$$

la cui regione ammissibile è rappresentata in Figura 4.5.

Nel punto $A = (1, 2)$ sono attivi i primi due vincoli, $g_1(x) = (x_1 - 1)^3 + (x_2 - 2) \leq 0$ e $g_2(x) = (x_1 - 1)^3 - (x_2 - 2) \leq 0$, i cui gradienti sono $\nabla g_1(A) = [0 \ 1]^T$ e $\nabla g_2(A) = [0 \ -1]^T$. La direzione $p = [1 \ 0]^T$ è ortogonale a entrambe, e quindi soddisfa le condizioni di ammissibilità. Nonostante ciò, nessun arco ammissibile ammette p come vettore tangente. Questo è possibile perché il punto A non è regolare. Infatti, i due gradienti $\nabla g_1(A)$ e $\nabla g_2(A)$ sono opposti, e quindi linearmente dipendenti. In termini intuitivi, descrivere i vincoli con i loro gradienti equivale a linearizzarli, cioè trattarli come rette, il che rende apparentemente possibile strisciare verso destra restando nella regione ammissibile.

Il caso dei vincoli di uguaglianza

C'è un'ulteriore complicazione, introdotta dai vincoli di uguaglianza, ma che si risolve facilmente introducendo un'eccezione. In generale, un vincolo di uguaglianza si può sostituire con due vincoli di disuguaglianza espressi da funzioni uguali e opposte:

$$h(x) = 0 \Leftrightarrow \begin{cases} h(x) \leq 0 \\ -h(x) \leq 0 \end{cases}$$

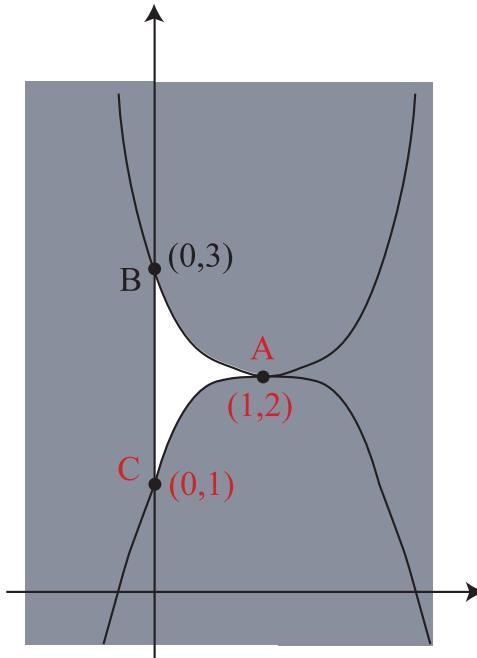


Figura 4.5: Regione ammissibile

La complicazione nasce dal fatto che, ovviamente, i gradienti di questi due vincoli sono anch'essi uguali e opposti ($\pm \nabla h(\tilde{x})$), e quindi linearmente dipendenti fra loro. Si potrebbe quindi concludere che nei problemi dotati di vincoli di uguaglianza tutti i punti sono non regolari. Fortunatamente, però, ogni singolo vincolo di uguaglianza, essendo $h(x) \in C^1(X)$, consente l'esistenza di archi ammissibili che partono da ogni punto ammissibile. Tali archi "strisciano" lungo il vincolo stesso, seguendolo rigorosamente. Solo la compresenza di altri vincoli può rendere inammissibili tali archi. Quindi, tutta la trattazione precedente rimane valida, con due precisazioni:

1. *nella definizione di punto regolare, per ogni vincolo di uguaglianza si considera solo uno dei due gradienti uguali e opposti;*
2. *nella definizione delle direzioni tangenti p_ξ , per ogni vincolo di uguaglianza si considerano entrambi i gradienti e poi si fondono le relative condizioni in una sola più semplice:*

$$\begin{cases} (\nabla h_i(\tilde{x}))^T p_\xi \leq 0 \\ -(\nabla h_i(\tilde{x}))^T p_\xi \leq 0 \end{cases} \Leftrightarrow (\nabla h_i(\tilde{x}))^T p_\xi = 0$$

cioè le direzioni p_ξ tangenti ad archi ammissibili sono ortogonali ai gradienti dei vincoli di uguaglianza.

Riscriviamo l'algoritmo

Ora possiamo riassumere le conclusioni raggiunte finora, prima di fare l'ultimo passaggio. Il procedimento che stiamo definendo è riportato in Figura 4.6 (si confronti con quello descritto in Figura 4.4) e consiste nei seguenti passaggi:

- si considerano tutti i punti ammissibili come candidati;

- si scorrono tutti i punti candidati;
- per ogni punto ammissibile \tilde{x} , si valuta la regolarità;
 - se \tilde{x} è non regolare, rimane candidato;
 - se \tilde{x} è regolare, si scorrono tutte le direzioni p che soddisfano le condizioni di ammissibilità sui gradienti dei vincoli attivi;
 - * per ognuna di tali direzioni p , si valuta la condizione sul gradiente della funzione obiettivo: se è violata, si scarta \tilde{x} ;
- si scorrono i punti candidati, conservandone i migliori.

```

Algorithm TrovaCandidati( $f, g_1, \dots, g_m$ )
 $C' := NonRegular(g);$ 
 $C := X \setminus C';$ 
For each  $x \in C$  do
  { Scarta i punti che violano le condizioni }
  For each  $p \in \mathbb{R}^n : (\nabla g_j(x))^T p \geq 0 \forall j : g_j(x) = 0$  do
    If  $(\nabla f(x))^T p < 0$  then  $C := C \setminus \{x\};$ 
  { Restituisce l'insieme dei punti candidati }
   $X^{KKT} := C \cup C';$ 
Return  $X^{KKT};$ 

```

Figura 4.6: Pseudocodice del procedimento per ridurre l'insieme ammissibile X al sottoinsieme dei punti candidati X_{KKT} .

Rimane solo da risolvere il problema che i due cicli su x e su p sono infiniti. Lo risolveremo sostituendo le relative condizioni con un sistema di equazioni e disequazioni, le cui soluzioni sono i punti candidati.

4.2.3 Una prima interpretazione geometrica

La proprietà del Corollario 1 ha un'interpretazione geometrica che la rende abbastanza intuitiva. Per ogni punto x preso in considerazione sono chiamati in causa due insiemi di vettori:

1. il *cono delle direzioni tangenti ad archi ammissibili* $C_{\text{amm}}(x)$ (in breve, *cono ammissibile*): è l'insieme dei vettori che formano angoli $\geq 90^\circ$ con i gradienti di tutti i vincoli attivi:

$$(\nabla g_j(x))^T p \leq 0 \text{ per ogni } j : g_j(x) = 0$$

Per ogni vincolo attivo, si ha un semispazio (quello che guarda “dalla parte opposta” del gradiente); nel complesso, bisogna intersecarli: l'insieme risultante è un cono.

2. il *semispazio migliorante* $C_{\text{migl}}(\tilde{x})$: è l'insieme dei vettori che formano angoli $< 90^\circ$ col gradiente della funzione obiettivo:

$$(\nabla f(x))^T p < 0$$

Il Corollario 1 si può esprimere dicendo che *il cono ammissibile non deve intersecare il semispazio migliorante*. Si noti che il primo è chiuso, mentre il secondo è aperto, cosicché i due insiemi si possono toccare, ma non possono avere punti interni comuni. Il motivo è che le direzioni in cui i due coni si toccano hanno prodotto scalare nullo col gradiente dell'obiettivo e nullo col gradiente di qualche vincolo attivo; in tal caso, le informazioni del primo ordine non sono sufficienti a determinare se la direzione sia migliorante o no, e se sia ammissibile o no. Quindi, tali punti non si possono scartare e devono rimanere candidati.

Esempio 15 *QUI VANNO ESEMPI TRATTI DAGLI ESERCIZI 1 E 5 PER ILLUSTRARE LA RELAZIONE FRA CONO AMMISSIBILE E SEMISPAZIO MIGLIORANTE*

4.3 Condizioni di Karush-Kuhn-Tucker

Questa sezione ribalta le condizioni del Corollario 1 in condizioni equivalenti, ma computazionalmente trattabili, cioè nella soluzione di un sistema di uguaglianze e disuguaglianze. Per farlo, si serve di un risultato teorico apparentemente molto lontano dai problemi fin qui trattati, che riguarda le proprietà di famiglie di vettori in spazi euclidei a molte dimensioni. Si tratta solo di mappare correttamente tali oggetti su quelli che stiamo manipolando. Per esprimere il lemma in poche parole, è necessario il concetto di combinazione conica di vettori, per il quale (come per altri concetti di algebra lineare) rimandiamo all'Appendice B.

4.3.1 Lemma di Farkas

Teorema 11 (*Lemma di Farkas*) *Siano $f \in \mathbb{R}^n$ e $g_j \in \mathbb{R}^n$ con $j = 1, \dots, m$ un vettore e una famiglia di vettori nello spazio euclideo a n dimensioni. Il vettore f è combinazione conica dei vettori g_j se e solo se:*

$$\exists \mu_j \geq 0 : f = \sum_{j=1}^m \mu_j g_j \Leftrightarrow p^T f \leq 0 \text{ per ogni } p : p^T g_j \leq 0 \text{ per } j = 1, \dots, m$$

cioè i vettori che puntano dalla parte opposta di tutti i g_j puntano anche dalla parte opposta di f .

Dimostrazione. Dimostriamo solo l'implicazione diretta di questo teorema (la parte “solo se”), che è banale: si tratta solo di sostituire f con $\sum_{j=1}^m \mu_j g_j$ e osservare che, quindi, $p^T f = p^T \sum_{j=1}^m \mu_j g_j = \sum_{j=1}^m \mu_j (p^T g_j)$. Qualsiasi vettore p per cui $p^T g_j \leq 0$ garantisce gli addendi della sommatoria siano non negativi, e quindi lo sia anche la sommatoria. Da qui la tesi.

L'implicazione inversa, invece, che è quella che ci serve, è più complessa, e non ne forniamo la dimostrazione. ■

Anche questo lemma ha un aspetto geometrico, che può renderlo più intuitivo. La Figura 4.7 fornisce un esempio con $m = 3$ vettori g_j in uno spazio a $n = 2$ dimensioni:

- per ogni vettore g_j , l'insieme dei vettori p che formano prodotti scalari $g_j^T p \leq 0$ è il semipiano opposto a g_j ;
- l'intersezione di tali semipiani è il cono (infinito) marcato in grigio in basso a sinistra;

- l'insieme dei vettori p che formano prodotti scalari $f^T p \leq 0$ è il semipiano opposto a f , che sta in basso a sinistra rispetto alla linea obliqua rossa;
- l'insieme dei vettori che sono combinazioni coniche dei vettori g_j è il cono (infinito) marcato in grigio in alto a destra.

Il lemma di Farkas dice che:

- se f cade nel cono dei vettori g , il cono opposto ai vettori g_j cade nel semipiano opposto a f ;
- se il cono opposto ai vettori g_j cade nel semipiano opposto a f , f cade nel cono dei vettori g .

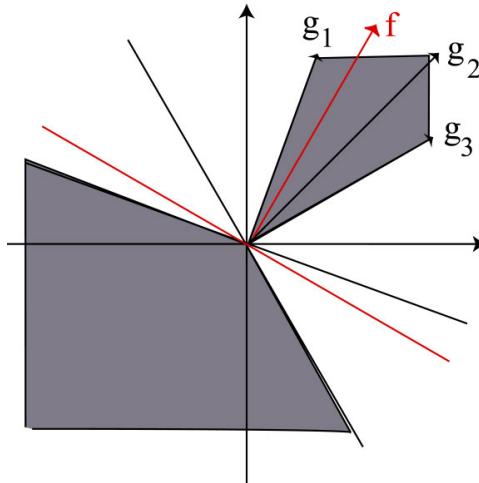


Figura 4.7: Esempio di applicazione del lemma di Farkas: f cade nel cono dei g_j se e solo se il cono opposto ai g_j cade nel semipiano opposto a f .

4.3.2 Forma standard delle condizioni di Karush-Kuhn-Tucker

A questo punto, applichiamo il lemma di Farkas ai vettori che stiamo utilizzando per filtrare i punti candidati. Precisamente, sostituiremo:

- i vettori g_j con i gradienti dei vincoli attivi $\nabla g_j(\tilde{x})$;
- il vettore f con l'antigradiente della funzione obiettivo $-\nabla f(\tilde{x})$.

Si noti il cambio di segno relativo al vettore f . Applicato a questi vettori, il lemma di Farkas dice che

$$\exists \mu_j \geq 0 : \nabla f(\tilde{x}) + \sum_{j: g_j(\tilde{x})=0} \mu_j \nabla g_j(\tilde{x}) = 0$$

se e solo se

$$(\nabla f(\tilde{x}))^T p \geq 0 \text{ per ogni } p : (\nabla g_j(\tilde{x}))^T p \leq 0 \text{ per ogni } j : g_j(\tilde{x}) = 0$$

Questo secondo membro significa che le direzioni ammissibili sono non miglioranti. Nei punti di ottimo locale regolari vale il secondo membro del lemma di Farkas;

quindi, vale anche il primo: se x^* è un punto regolare di ottimo locale, allora in esso l'antigradiente della funzione obiettivo cade nel cono dei gradienti dei vincoli attivi:

$$\exists \mu_j \geq 0 : \nabla f(x^*) + \sum_{j:g_j(x^*)=0} \mu_j \nabla g_j(x^*) = 0$$

In genere, le condizioni di KKT si presentano in una forma equivalente, con due piccole modifiche. Per prima cosa, si può osservare che la combinazione conica dei gradienti si può estendere dai soli vincoli attivi a tutti i vincoli a patto di imporre moltiplicatori nulli ai vincoli inattivi. Questo si può ottenere introducendo delle condizioni aggiuntive che legano i moltiplicatori e i vincoli, le così dette *condizioni di complementarietà*:

$$\sum_{j:g_j(x^*)=0} \mu_j \nabla g_j(x^*) = \sum_{j=1}^s \mu_j \nabla g_j(x^*) \text{ con } \mu_j g_j(x) = 0, \forall j$$

Inoltre, in genere i vincoli di uguaglianza non vengono scomposti in vincoli di disuguaglianza, ma si osserva che, anziché due gradienti opposti con moltiplicatori $\mu^+ \geq 0$ e $\mu^- \geq 0$, si può usare un solo gradiente con moltiplicatore pari alla differenza dei due, cioè un moltiplicatore libero in segno:

$$h_i(x) = 0 \Leftrightarrow \begin{cases} g_{j'_i}(x) = h_i(x) & \leq 0 \\ g_{j''_i}(x) = -h_i(x) & \leq 0 \end{cases}$$

$$\begin{aligned} \dots + \mu_{j'_i} \nabla h_i(x) + \mu_{j''_i} (-\nabla h_i(x)) + \dots & \quad \text{con } \mu_{j'_i}, \mu_{j''_i} \geq 0 \\ \dots + \lambda_i \nabla h_i(x) + \dots & \quad \text{con } \lambda_i = \mu_{j'_i} - \mu_{j''_i} \text{ libero} \end{aligned}$$

Teorema 12 (Condizioni di KKT) *Sia $X = \{x \in \mathbb{R}^n : h_i(x) = 0, g_j(x) \leq 0, \text{ con } i = 1, \dots, s \text{ e } j = 1, \dots, m\}$ e siano $f, h_i, g_j \in C^1(X)$ per $i = 1, \dots, s$ e $j = 1, \dots, m$. Se x^* è un punto regolare in X e un punto di minimo locale per f in X , allora esistono moltiplicatori λ_i liberi in segno e $\mu_j \geq 0$ tali che:*

$$\nabla f(x^*) + \sum_{i=1}^s \lambda_i \nabla h_i(x^*) + \sum_{j=1}^m \mu_j \nabla g_j(x^*) = 0 \tag{4.2a}$$

$$h_i(x^*) = 0 \quad i = 1, \dots, s \tag{4.2b}$$

$$\mu_j g_j(x^*) = 0 \quad j = 1, \dots, m \tag{4.2c}$$

$$g_j(x^*) \leq 0 \quad j = 1, \dots, m \tag{4.2d}$$

$$\mu_j \geq 0 \quad j = 1, \dots, m \tag{4.2e}$$

Questo significa che l'algoritmo di filtraggio non consiste nello scorrere tutti i punti $x \in X$ e le direzioni $p \in \mathbb{R}^n$, valutando se soddisfino opportune condizioni, ma consiste nel risolvere un sistema di equazioni e disequazioni, determinando i punti che lo soddisfano. Si può osservare che il sistema lega $n + s + m$ variabili (x, λ, μ) con $n + s + m$ uguaglianze (4.2a, 4.2b, 4.2c), oltre a $2m$ disuguaglianze (4.2d, 4.2e). Il sistema, quindi, è bilanciato, e ci si può aspettare che abbia in generale un numero di soluzioni finito (non una sola, perché non è lineare).

Concludendo, le condizioni in forma standard sono anche a volte descritte introducendo una funzione ausiliaria, detta *funzione lagrangiana generalizzata*

$$\ell(x, \lambda, \mu) = f(x) + \sum_{i=1}^s \lambda_i h_i(x) + \sum_{j=1}^m \mu_j g_j(x)$$

Data questa definizione, le Equazioni (4.2) si possono interpretare come l'azzeramento delle derivate parziali di $\ell(x, \lambda, \mu)$ rispetto alle variabili x , mentre le Equazioni (4.2b) come l'azzeramento delle derivate parziali di $\ell(x, \lambda, \mu)$ rispetto alle variabili *lambda*. Le condizioni di complementarietà (4.2d), e le condizioni sul segno dei moltiplicatori (4.2e), però, non hanno un'intepretazione legata alla lagrangiana generalizzata.

4.3.3 Una seconda interpretazione geometrica

Dal punto di vista geometrico, l'antigradiente della funzione obiettivo, $-\nabla f(x)$, è la direzione in cui l'obiettivo f cala più rapidamente, cioè la direzione più migliorante. I gradienti dei vincoli attivi, $\nabla g_j(x^*)$, sono le direzioni in cui le funzioni g_j crescono più rapidamente, cioè le direzioni di massima violazione dei vincoli. Quando l'antigradiente dell'obiettivo cade nel cono dei gradienti dei vincoli attivi, l'obiettivo può migliorare solo violando almeno un vincolo.

Esempio 16 ESEMPIO BASATO SULL'ESERCIZIO 1: SI VEDE UN PUNTO GENERICO, UN OTTIMO LOCALE E UN PUNTO CANDIDATO CHE NON E' OTTIMO LOCALE

Casi particolari interessanti

Vi sono diversi casi particolari che risultano interessanti per vari motivi.

Problemi non vincolati Anzi tutto, c'è il caso dei problemi non vincolati. Se non ci sono vincoli, le condizioni di KKT si riducono a richiedere l'annullamento del gradiente dell'obiettivo:

$$\nabla f(x^*) = 0$$

Questa condizione non è altro che la generalizzazione della condizione di annullamento della derivata prima per calcolare i minimi di funzioni a una sola variabile. I punti candidati così determinati possono essere minimi globali, semplicemente locali, ma anche punti di flesso e persino punti di massimo locale.

Problemi discreti

Infine, c'è il caso, vedremo piuttosto sfortunato, dei problemi discreti, nei quali lo spazio ammissibile è composto da punti isolati (per esempio a coordinate intere: $X \subseteq \mathbb{Z}^n$). Si potrebbe pensare di aggirare l'ostacolo esprimendo il vincolo di interezza con un'espressione regolare come $h_i(x) = \sin(\pi x_i) = 0$. Con questa ipotesi (o qualsiasi altra equivalente), le condizioni di KKT si possono applicare, ma risultano praticamente inutili. Infatti, i vincoli $h_i(x) = 0$ introducono in ogni equazione del sistema un termine aggiuntivo $\lambda_i \frac{\partial h_i}{\partial x_i}$, che compare solo in tale equazione. Questo significa che, qualunque punto x si consideri, è possibile assegnare a ogni moltiplicatore λ_i un valore tale da soddisfare l'equazione corrispondente: *nei problemi discreti, tutti i punti sono candidati*. Questo è intuitivamente ovvio, dato che nei problemi discreti ogni punto è isolato, cioè ammette un intorno abbastanza piccolo da essere l'unico punto ammissibile in esso contenuto. Quindi, in un problema discreto ogni punto è di ottimo locale.

4.4 Applicazioni delle condizioni di KKT*

4.4.1 Pianificazione per sistemi Big-Data

I sistemi di gestione Big-Data come Hadoop/YARN gestiscono computazioni con caratteristiche e livelli di servizio richiesti molto eterogenei. Questo complica il compito di decidere a quante e a quali risorse assegnare le richieste di calcolo e come distribuire nel tempo le operazioni, ottenendo insieme un alto livello di utilizzo delle risorse e una buona soddisfazione delle richieste di calcolo. Nel seguito viene descritto un piccolo sottoproblema di gestione nel quale si possono direttamente applicare le condizioni di KKT⁴.

È dato un insieme J di $n = |J|$ operazioni da eseguire (*job*). Ogni *job* $j \in J$ è associato a un carico di lavoro complessivo w_j , che può essere suddiviso tra più risorse in modo che il lavoro proceda in parallelo. Assumeremo i *job* perfettamente *malleabili*, cioè divisibili del tutto a piacere in ogni istante di tempo, a patto di rispettare una durata minima e una durata massima (t_j^{\min} e t_j^{\max}), nonché una quantità minima e massima di risorse utilizzabili in parallelo (r_j^{\min} e r_j^{\max}). Ovviamente, si tratta di un'approssimazione, che può essere giustificata per *job* pesantemente parallelizzabili. I *job* di J vanno eseguiti in una sequenza preassegnata, quindi formano una catena completamente ordinata. Si vuole eseguire la catena di *job* minimizzando la variazione totale di utilizzo delle risorse durante l'orizzonte temporale T . La variazione di utilizzo è definita in ogni istante come il modulo della variazione nella quantità di risorse utilizzata rispetto all'istante precedente. L'idea è che, anche se trascuriamo il tempo necessario per attivare e disattivare le risorse, queste operazioni hanno un costo che si vuole minimizzare.

Possiamo rappresentare i *job* come figure geometriche bidimensionali, dove l'asse orizzontale corrisponde al tempo, mentre quello verticale corrisponde alla quantità di risorse impiegata in ciascun istante. È abbastanza intuitivo che la soluzione migliore dal punto di vista della variazione di risorse corrisponde a una figura rettangolare, cioè il *job* j viene diviso fra r_j risorse che lo eseguono per t_j unità di tempo, dopo di che il *job* termina. Questo perché, fissato il carico di lavoro, cioè l'area della figura, far crescere e calare più volte l'altezza, cioè le risorse impegnate, produrrebbe solo una variazione maggiore. D'altra parte, farle crescere e calare lentamente produrrebbe solo una dilatazione del tempo impiegato a eseguire l'intero *job*.

A questo punto, vogliamo trovare le basi e le altezze dei rettangoli in J in modo da minimizzare la somma delle loro altezze (contate due volte, perché la variazione della quantità di risorse impiegate avviene sia al principio sia al termine del *job*). I rettangoli vanno posti in sequenza nel tempo (asse orizzontale) e l'ampiezza totale

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

⁴Questa applicazione è tratta da R. Cordone, G. M. Fumarola, M. Mazzucchelli, M. Rabozzi, M. Santambrogio, *Preemption-aware planning on Big-Data Systems*, PPoPP '16 Proceedings of the 21st ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, Article No. 48.

della catena non deve superare L . Ne deriva il seguente modello:

$$\begin{aligned} \min_{r,t} f(r,t) &= \sum_{j \in J} 2r_j \\ r_j t_j &= w_j \quad j \in J \\ \sum_{j \in J} t_j &\leq T \\ r_j^{\min} &\leq r_j \leq r_j^{\max} \quad j \in J \\ t_j^{\min} &\leq t_j \leq t_j^{\max} \quad j \in J \end{aligned}$$

Il problema ha $2|J|$ variabili, $4|J| + 1$ vincoli di disuguaglianza e $|J|$ vincoli di uguaglianza.

L'applicazione diretta delle condizioni di KKT produce un sistema tutt'altro che banale. È però possibile fare qualche semplificazione:

1. dividiamo per 2 la funzione obiettivo, il che non modifica la soluzione ottima;
2. usiamo i vincoli $r_j t_j = w_j$ per ricavare le variabili r_j in funzione delle variabili t_j ; questo elimina $|J|$ variabili e i $|J|$ vincoli di uguaglianza;
3. osserviamo che i vincoli $r_j^{\min} \leq r_j \leq r_j^{\max}$, a valle della sostituzione diventano $w_j/r_j^{\max} \leq t_j \leq w_j/r_j^{\min}$; si possono combinare questi nuovi vincoli su t_j con i vecchi, scegliendo caso per caso il vincolo più stretto.

Si ottiene quindi la formulazione semplificata:

$$\begin{aligned} \min f(t) &= \sum_{j \in J} \frac{w_j}{t_j} \\ \sum_{j \in J} t_j &\leq T \\ \bar{t}_j^{\min} &\leq t_j \leq \bar{t}_j^{\max} \quad j \in J \end{aligned}$$

dove $\bar{t}_j^{\min} = \max(t_j^{\min}, w_j/r_j^{\max})$ e $\bar{t}_j^{\max} = \min(t_j^{\max}, w_j/r_j^{\min})$. Questa formulazione ha $|J|$ variabili e $2|J| + 1$ vincoli di disuguaglianza.

Assegnato un moltiplicatore $\mu \geq 0$ al vincolo $\sum_{j \in J} t_j \leq T$ e moltiplicatori $\nu_j^{\min} \geq 0$ e $\nu_j^{\max} \geq 0$ ai vincoli che impongono i valori estremi di t_j , si ottiene il seguente sistema di condizioni di KKT:

$$\begin{aligned} -2 \frac{w_j}{t_j^2} + \mu - \nu_j^{\min} + \nu_j^{\max} &= 0 \quad j \in J \\ \mu \left(\sum_{j \in J} t_j - T \right) &= 0 \\ \nu_j^{\min} (t_j - \bar{t}_j^{\min}) &= 0 \quad j \in J \\ \nu_j^{\max} (t_j - \bar{t}_j^{\max}) &= 0 \quad j \in J \\ \sum_{j \in J} t_j &\leq T \\ \bar{t}_j^{\min} &\leq t_j \leq \bar{t}_j^{\max} \quad j \in J \\ \mu, \nu_j^{\min}, \nu_j^{\max} &\geq 0 \quad j \in J \end{aligned}$$

che è ancora non elementare.

Ora dovremmo considerare le condizioni di complementarietà e dividere il problema in sottoproblemi. I sottoproblemi corrispondono a regioni diverse dello spazio delle soluzioni, nelle quali diverse famiglie di vincoli sono attive o inattive.

Faremo invece una cosa molto empirica: analizzeremo uno solo dei sottoproblemi, che appare il più plausibile. La maggior parte dei vincoli di disuguaglianza impone alle durate dei *job*, t_j , di cadere in un dato intervallo $[t_j^{\min}; t_j^{\max}]$. Se i valori t_j^{\min} sono sufficientemente piccoli e i valori t_j^{\max} sono sufficientemente grandi, cioè se i *job* sono sufficientemente malleabili, è plausibile che nella soluzione ottima questi vincoli siano tutti inattivi. Infatti, *job* con durate t_j brevi implicano un uso di risorse r_j alto, che impatta sull'obiettivo pesantemente. D'altra parte, *job* con durate t_j lunghe consumano pesantemente il tempo T a disposizione, e quindi impattano sul vincolo $\sum_{j \in J} t_j \leq T$, che viene conservato. Riassumendo, invece di analizzare tutte le $2|J| + 1$ condizioni di complementarietà, producendo (nel caso peggiore) $2^{2|J|+1}$ sottoproblemi, analizzeremo solo una condizione. Il rischio è di perdere soluzioni candidate, eventualmente ottime.

Limitandosi quindi ai sottoproblemi in cui $v_j^{\min} = v_j^{\max} = 0$ per ogni $j \in J$, il sistema semplificato diventa:

$$\begin{aligned} -\frac{w_j}{t_j^2} + \mu &= 0 & j \in J \\ \mu \left(\sum_{j \in J} t_j - T \right) &= 0 \\ \sum_{j \in J} t_j &\leq T \\ \bar{t}_j^{\min} \leq t_j &\leq \bar{t}_j^{\max} & j \in J \\ \mu &\geq 0 & j \in J \end{aligned}$$

Il primo vincolo impone $\mu = 2w_j/t_j^2 > 0$, e quindi $\sum_{j \in J} t_j = T$: l'orizzonte temporale viene interamente sfruttato (cosa abbastanza intuitiva, perché permette di usare meno risorse, e quindi risparmiare sulle variazioni delle risorse impiegate). Risolvendo rispetto ai t_j , ricaviamo il loro valore rispetto a μ :

$$\mu = \frac{w_j}{t_j^2} \Rightarrow t_j^2 = \frac{w_j}{\mu} \Rightarrow t_j = \sqrt{\frac{w_j}{\mu}}$$

A questo punto, il vincolo sull'orizzonte temporale fornisce il valore di μ :

$$\sum_{j \in J} t_j = \sum_{j \in J} \sqrt{\frac{w_j}{\mu}} = T \Rightarrow \sqrt{\mu} = \frac{\sum_{j \in J} \sqrt{w_j}}{T}$$

da cui

$$t_j = \frac{\sqrt{w_j}}{\sum_{j \in J} \sqrt{w_j}} T \quad \text{per ogni } j \in J$$

che è una soluzione molto elegante, dato che distribuisce l'orizzonte temporale T fra i vari *job* in proporzione non al carico di lavoro (che è un'“area”), ma alla sua radice quadrata (che è una “lunghezza”).

Il valore corrispondente dell'obiettivo è:

$$f^* = \sum_{j \in J} 2r_j^* = 2 \sum_{j \in J} \frac{w_j}{t_j^*} = 2T \sum_{j \in J} \sqrt{w_j} \sum_{j \in J} \sqrt{w_j} = 2T \left(\sum_{j \in J} \sqrt{w_j} \right)^2$$

Questa soluzione è valida se ogni t_j rimane compreso fra t_j^{\min} e t_j^{\max} . Se anche solo uno di tali valori esce dal relativo intervallo, la soluzione va scartata. Inoltre, potrebbero esistere soluzioni, migliori o peggiori, negli altri sottoproblemi: per saperlo dovremmo analizzarli tutti. In realtà, si può dimostrare con condizioni del secondo ordine (che in queste note non trattiamo) che il problema, essendo convesso, ammette una sola soluzione candidata, che è globalmente ottima. Quindi, se la soluzione trovata è ammissibile, è anche globalmente ottima. Se non lo è, esistono algoritmi efficienti (che non trattiamo qui) per estendere l'algoritmo del simplexso dalla programmazione lineare ai problemi con funzione obiettivo convessa e vincoli lineari, in modo da trovare la soluzione globalmente ottima senza esplorare tutti i sottoproblemi.

4.5 Esercizi

Esercizio 1

Si vuole localizzare una discarica di rifiuti nucleari il più vicino possibile alla centrale che li produce (posizione $\bar{x} = (1, 0)$). D'altra parte, la discarica deve stare a una distanza pari almeno a 2 da una città posta in $\tilde{x} = (0, 0)$ e deve stare fuori da un parco: la linea di rispetto corre lungo la retta $x = 3/2$ e la centrale deve stare sul versante sinistro della retta stessa.

Soluzione

Modello

Detta $x = (x_1, x_2)$ la posizione della discarica, il problema ammette il seguente modello di programmazione matematica:

$$\begin{aligned}\min f(x) &= (x_1 - 1)^2 + x_2^2 \\ g_1(x) &= -x_1^2 - x_2^2 + 4 \leq 0 \\ g_2(x) &= x_1 - 3/2 \leq 0\end{aligned}$$

La Figura 4.8 rappresenta la regione delle soluzioni ammissibili.

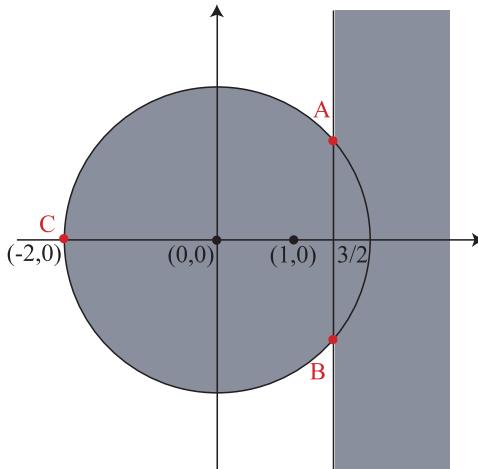


Figura 4.8: Regione ammissibile

Punti non regolari

I punti regolari sono quei punti nei quali i gradienti dei vincoli attivi sono fra loro linearmente indipendenti. Si noti che vanno considerati solo i vincoli attivi. Quindi, i punti interni alla regione di ammissibilità sono tutti regolari, quelli per i quali passa un vincolo solo lo sono ovunque il gradiente è diverso da zero, e così via.

I gradienti dei due vincoli sono:

$$\begin{aligned}\nabla g_1(x) &= [-2x_1 \ -2x_2]' \\ \nabla g_2(x) &= [1 \ 0]'\end{aligned}$$

Il primo gradiente si annulla solo nell'origine, dove però il vincolo non è attivo ($g_1(0,0) = 4 > 0$). Il secondo gradiente non si annulla mai. Ci interessano quindi solo i punti in cui sono attivi entrambi i vincoli (indicati con A e B nella figura):

$$\begin{cases} g_1(x) = -x_1^2 - x_2^2 + 4 = 0 \\ g_2(x) = x_1 - 3/2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 3/2 \\ x_2^2 = 4 - x_1^2 = 7/4 \end{cases}$$

da cui $A = (3/2, \sqrt{7}/2)$ e $B = (3/2, -\sqrt{7}/2)$.

In A è $\nabla g_1(x) = [-3 \ -\sqrt{7}]'$ e i gradienti dei due vincoli sono linearmente indipendenti. Questo si può verificare in due modi: con la definizione o accostando i vettori per costruire una matrice di cui si valuta il rango. Il primo sistema consiste nell'imporre

$$\alpha_1 \nabla g_1(A) + \alpha_2 \nabla g_2(A) = 0$$

da cui

$$\begin{cases} -3\alpha_1 + \alpha_2 = 0 \\ -\sqrt{7}\alpha_1 = 0 \end{cases} \Rightarrow \alpha_1 = \alpha_2 = 0$$

Poiché l'unico modo per annullare una combinazione lineare dei due vettori è di usare coefficienti nulli, i due vettori sono linearmente indipendenti per definizione.

D'altra parte, la matrice

$$\begin{bmatrix} -3 & 1 \\ -\sqrt{7} & 0 \end{bmatrix}$$

ha rango 2, poiché il suo determinante è non nullo: $-3 \cdot 0 + \sqrt{7} \cdot 1 = \sqrt{7} \neq 0$

Condizioni di Karush-Kuhn-Tucker

Le condizioni di Karush-Kuhn-Tucker (anche note come condizioni necessarie del primo ordine) affermano che se un punto è regolare e di minimo locale, allora

1. sono nulle le derivate parziali della lagrangiana generalizzata rispetto alle variabili x ($\partial \ell / \partial x_i = 0$)
2. sono nulle le derivate parziali della lagrangiana generalizzata rispetto ai moltiplicatori λ ($\partial \ell / \partial \lambda_j = h_j = 0$), cioè sono rispettati i vincoli di uguaglianza
3. sono nulli i prodotti delle funzioni che esprimono i vincoli di disuguaglianza per i loro moltiplicatori ($\mu_k g_k = 0$)
4. sono soddisfatti tutti i vincoli di disuguaglianza ($g_k \leq 0$)
5. sono non negativi i moltiplicatori dei vincoli di disuguaglianza ($\mu_k \geq 0$)

Usiamo queste condizioni per eliminare i punti che non le soddisfano, nella speranza che rimangano solo pochi *punti candidati*. A tali punti dovremo aggiungere i punti non regolari, dato che in essi le condizioni di K.K.T. non sono necessarie. In questo caso, però, tutti i punti sono regolari.

La lagrangiana generalizzata è $\ell(x) = f(x) + \mu_1 g_1(x) + \mu_2 g_2(x) = (x_1 - 1)^2 + x_2^2 + \mu_1(-x_1^2 - x_2^2 + 4) + \mu_2(x_1 - 3/2)$, per cui

$$\begin{aligned}\partial\ell/\partial x_1 &= 2(x_1 - 1) - 2\mu_1 x_1 + \mu_2 = 0 \\ \partial\ell/\partial x_2 &= 2x_2 - 2\mu_1 x_2 = 0 \\ \mu_1 g_1 &= \mu_1(-x_1^2 - x_2^2 + 4) = 0 \\ \mu_2 g_2 &= \mu_2(x_1 - 3/2) = 0 \\ g_1 &= -x_1^2 - x_2^2 + 4 \leq 0 \\ g_2 &= x_1 - 3/2 \leq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0\end{aligned}$$

Per risolvere questo sistema senza esplorare esaustivamente tutti i casi possibili, si può impostare un albero di ricerca, che ad ogni nodo divida la regione ammissibile in parti. Per farlo, ci basiamo sulle condizioni più fastidiose, che sono i prodotti $\mu_k g_k = 0$. Scelto un vincolo, possiamo distinguere tre casi:

1. $\mu_k = 0$ e $g_k < 0$
2. $\mu_k = 0$ e $g_k = 0$
3. $\mu_k > 0$ e $g_k = 0$

ma in genere i calcoli risultano molto più semplici se si include il secondo caso nel primo invece di trattarlo a parte o di includerlo nel terzo. Si pone quindi:

1. $\mu_k = 0$ e $g_k \leq 0$
2. $\mu_k > 0$ e $g_k = 0$

Quanto al vincolo, conviene scegliere quello più semplice, cioè il secondo e spezzare il problema originario P^0 in P^1 ($\mu_2 = 0$ e $g_2 \leq 0$) e P^2 ($\mu_2 > 0$ e $g_2 = 0$).

P¹ ($\mu_2 = 0$ e $g_2 \leq 0$)

$$\begin{aligned}x_1(1 - \mu_1) &= 1 \\ x_2(1 - \mu_1) &= 0 \\ \mu_1(-x_1^2 - x_2^2 + 4) &= 0 \\ x_1^2 + x_2^2 &\geq 4 \\ x_1 &\leq 3/2 \\ \mu_1 &\geq 0\end{aligned}$$

Il primo vincolo garantisce che $\mu_1 \neq 1$, per cui $x_2 = 0$

$$\begin{aligned}x_1(1 - \mu_1) &= 1 \\ x_2 &= 0 \\ \mu_1(4 - x_1^2) &= 0 \\ x_1^2 &\geq 4 \\ x_1 &\leq 3/2 \\ \mu_1 &\geq 0\end{aligned}$$

Poiché $x_1^2 \geq 4$ e $x_1 \leq 3/2$, è $x_1 \leq -2$ e quindi $1 - \mu_1 = 1/x_1 < 0$, per cui $\mu_1 > 1$. Quindi $4 - x_1^2 = 0 \Rightarrow x_1 = -2$. Se ne ricava il punto candidato $C = (-2, 0)$ con $\mu_C = (3/2, 0)$.

P² ($\mu_2 > 0$ e $g_2(x) = 0 \Rightarrow x_1 = 3/2$)

$$\begin{aligned} 1 - 3\mu_1 + \mu_2 &= 0 \\ x_2(1 - \mu_1) &= 0 \\ \mu_1(7/4 - x_2^2) &= 0 \\ 7/4 - x_2^2 &\leq 0 \\ \mu_1 &\geq 0 \end{aligned}$$

Poiché $\mu_1 = (\mu_2 + 1)/3 \geq 1/3 > 0$, è $x_2^2 = 7/4$. Ne derivano i due punti candidati: $A = (3/2, \sqrt{7}/2)$ con $\mu_A = (1, 2)$ e $B = (3/2, -\sqrt{7}/2)$ con $\mu_B = (1, 2)$.

Abbiamo quindi in complesso tre punti candidati, fra i quali scegliere quello in cui la funzione obiettivo assume il valore minimo.

$$\begin{cases} A = (3/2, \sqrt{7}/2) \Rightarrow f(A) = 2 \\ B = (3/2, -\sqrt{7}/2) \Rightarrow f(B) = 2 \\ C = (-2, 0) \Rightarrow f(C) = 9 \end{cases}$$

da cui risulta che A e B sono entrambi punti di ottimo globale.

Significato geometrico delle condizioni

È particolarmente istruttivo osservare nei punti candidati e in altri ancora i gradienti dei vincoli attivi e il gradiente della funzione obiettivo.

In A è

$$\nabla g_1(A) = \begin{bmatrix} -3 & -\sqrt{7} \end{bmatrix}' \quad \nabla g_2 \equiv [1 \ 0]' \quad \nabla f(A) \equiv \begin{bmatrix} 1 & \sqrt{7} \end{bmatrix}'$$

Ogni direzione ammissibile (cioè che forma angoli $\geq 90^\circ$ con i gradienti dei vincoli attivi, $\nabla g_1(A)$ e $\nabla g_2(A)$) forma angoli $\leq 90^\circ$ anche con $\nabla f(A)$, ed è quindi non migliorante. Il lemma di Farkas garantisce che in tal caso, il vettore $-\nabla f(A)$ cade nel cono formato dai vettori $\nabla g_1(A)$ e $\nabla g_2(A)$. Infatti $-\nabla f(A) = \nabla g_1(A) + 2\nabla g_2(A)$.

In B succedono le stesse cose, semplicemente ribaltate rispetto all'asse x_1 .

In C è attivo solo g_1 , per cui

$$\nabla g_1(C) = [4 \ 0]' \quad \nabla f(C) \equiv [-6 \ 0]'$$

Le direzioni ammissibili formano un cono degenere che è in realtà il semispazio rivolto verso sinistra. Tutte formano angoli acuti con la direzione del gradiente e sono quindi non miglioranti. Obbedendo al lemma di Farkas, il vettore $-\nabla f(A)$ cade nel cono degenere (una semiretta) formato dal gradiente dell'unico vincolo attivo $\nabla g_1(C)$. Infatti $-\nabla f(A) = 3/2 \nabla g_1(A)$.

In $(0, 2)$

$$\nabla g_1(0, 2) = [0 \ -4]' \quad \nabla f(0, 2) \equiv [-2 \ 4]'$$

per cui le direzioni ammissibili sono quelle rivolte verso l'alto, e costituiscono ancora un cono degenere. La direzione del gradiente è rivolta verso l'alto, ma un po' a sinistra, e quindi forma un angolo ottuso con alcune direzioni ammissibili. Tali direzioni risultano anche migliori. Obbedendo al lemma di Farkas, l'opposto del gradiente non cade nel cono (degenero) individuato da $\nabla g_1(0, 2)$.

In $(-2, 2)$

$$\nabla f(-2, 2) \equiv [-4 \ 4]'$$

e non vi sono vincoli attivi. Tutte le direzioni sono ammissibili, e alcune di esse formano angoli ottusi con $\nabla f(-2, 2)$. D'altra parte, il cono individuato dai vincoli attivi è vuoto, dato che non vi sono vincoli attivi. Secondo il lemma di Farkas, l'opposto del gradiente non può cadere in tale cono. E in effetti, potrebbe solo se si annullasse, ma non è nullo.

Esercizio 2

Risolvere con le condizioni di KKT il seguente problema:

$$\begin{aligned} \min f(x) &= x_1 + x_2 \\ g_1(x) &= -x_1^2 + x_2^2 + 1 \leq 0 \\ g_2(x) &= x_1^2 - 4 \leq 0 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

Soluzione

La Figura 4.9 rappresenta la regione delle soluzioni ammissibili.

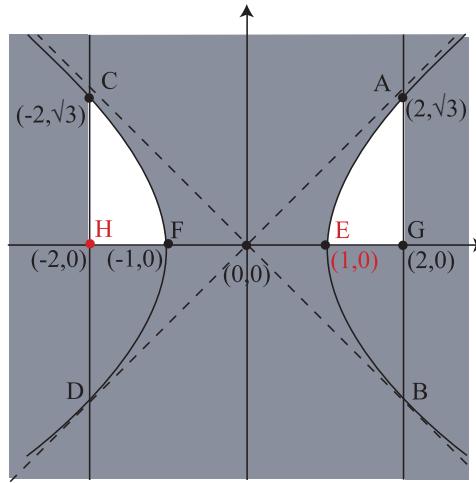


Figura 4.9: Regione ammissibile

Punti non regolari

I punti regolari sono quei punti nei quali i gradienti dei vincoli attivi sono fra loro linearmente indipendenti. Si noti che vanno considerati solo i vincoli attivi. Quindi, i punti interni alla regione di ammissibilità sono tutti regolari, quelli per i quali passa un vincolo solo lo sono ovunque il gradiente è diverso da zero, e così via.

I gradienti dei tre vincoli sono:

$$\begin{aligned} \nabla g_1(x) &= [-2x_1 \ 2x_2]' \\ \nabla g_2(x) &= [2x_1 \ 0]' \\ \nabla g_3(x) &= [0 \ -1]' \end{aligned}$$

Il primo e il secondo gradiente si annullano solo nell'origine, dove i relativi vincoli non sono attivi. Le intersezioni dei vincoli a coppie sono:

- $g_1(x) = g_2(x) = 0$

$$\begin{aligned} g_1(x) &= -x_1^2 + x_2^2 + 1 \leq 0 \\ g_2(x) &= x_1^2 - 4 \leq 0 \end{aligned}$$

da cui si ricavano i punti $A = (2, \sqrt{3})$, $B = (2, -\sqrt{3})$, $C = (-2, \sqrt{3})$ e $D = (-2, -\sqrt{3})$. I gradienti ∇g_1 e ∇g_2 valutati in questi quattro punti sono linearmente indipendenti. Verifichiamolo per A (il procedimento è analogo per gli altri tre punti):

$$\nabla g_1 = \begin{bmatrix} -4 & 2\sqrt{3} \end{bmatrix}' \quad \nabla g_2 = [4 \ 0]'$$

che sono linearmente indipendenti perché la matrice ottenuta accostandoli è non singolare:

$$\begin{bmatrix} -4 & 4 \\ 2\sqrt{3} & 0 \end{bmatrix} \Rightarrow \begin{vmatrix} -4 & 4 \\ 2\sqrt{3} & 0 \end{vmatrix} = -8\sqrt{3} \neq 0$$

- $g_1(x) = g_3(x) = 0$

$$\begin{aligned} g_1(x) &= -x_1^2 + x_2^2 + 1 \leq 0 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

da cui si ricavano i punti $E = (1, 0)$ ed $F = (-1, 0)$. Anche i gradienti ∇g_1 e ∇g_3 valutati in questi due punti sono linearmente indipendenti.

- $g_2(x) = g_2(x) = 0$

$$\begin{aligned} g_2(x) &= x_1^2 - 4 \leq 0 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

da cui si ricavano i punti $G = (2, 0)$ e $H = (-2, 0)$, in cui i gradienti ∇g_2 e ∇g_3 sono linearmente indipendenti.

Infine, non vi sono intersezioni di tutti e tre i vincoli. Quindi, tutti i punti sono regolari.

Condizioni di Karush-Kuhn-Tucker

La lagrangiana generalizzata è $\ell(x, \mu) = x_1 + x_2 + \mu_1(-x_1^2 + x_2^2 + 1) + \mu_2(x_1^2 - 4) + \mu_3(-x_2)$. Quindi le condizioni di KKT sono:

$$\begin{aligned} \partial \ell / \partial x_1 &= 1 - 2\mu_1 x_1 + 2\mu_2 x_1 = 0 \\ \partial \ell / \partial x_2 &= 1 + 2\mu_1 x_2 - \mu_3 = 0 \\ \mu_1 g_1 &= \mu_1(-x_1^2 + x_2^2 + 1) = 0 \\ \mu_2 g_2 &= \mu_2(x_1^2 - 4) = 0 \\ \mu_3 g_3 &= \mu_3 x_2 = 0 \\ g_1 &= -x_1^2 + x_2^2 + 1 \leq 0 \\ g_2 &= x_1^2 - 4 \leq 0 \\ g_3 &= -x_2 \leq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \\ \mu_3 &\geq 0 \end{aligned}$$

Il vincolo più semplice è $\mu_3 g_3(x) = 0$, per cui spezziamo il problema originario P^0 in P^1 ($\mu_3 = 0$ e $g_3 \leq 0$) e P^2 ($\mu_3 > 0$ e $g_3 = 0$).

P¹ ($\mu_3 = 0$ e $g_3 \leq 0$)

$$\begin{aligned} 2(\mu_1 - \mu_2)x_1 &= 1 \\ 1 + 2\mu_1x_2 &= 0 \\ \mu_1(-x_1^2 + x_2^2 + 1) &= 0 \\ \mu_2(x_1^2 - 4) &= 0 \\ -x_1^2 + x_2^2 + 1 &\leq 0 \\ x_1^2 &\leq 4 \\ x_2 &\geq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \end{aligned}$$

Poiché $\mu_1 \geq 0$ e $x_2 \geq 0$, è $1 + 2\mu_1x_2 \geq 1 > 0$. Quindi il secondo vincolo non può essere soddisfatto e questo sottoproblema non contiene punti candidati.

P² ($\mu_3 > 0$ e $g_3 = 0 \Rightarrow x_2 = 0$)

$$\begin{aligned} 2(\mu_1 - \mu_2)x_1 &= 1 \\ 1 - \mu_3 &= 0 \Rightarrow \mu_3 = 1 \\ \mu_1(-x_1^2 + 1) &= 0 \\ \mu_2(x_1^2 - 4) &= 0 \\ -x_1^2 + 1 &\leq 0 \\ x_1^2 - 4 &\leq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \end{aligned}$$

Scomponiamo ulteriormente il sottoproblema P^2 , rispetto al vincolo $\mu_1 g_1(x) = 0$, ottenendo P^3 ($\mu_1 = 0$ e $g_1 \leq 0$) e P^4 ($\mu_1 > 0$ e $g_1 = 0$).

P³ ($\mu_1 = 0$ e $g_1 \leq 0$)

$$\begin{aligned} -2\mu_2x_1 &= 1 \\ \mu_3 &= 1 \\ \mu_2(x_1^2 - 4) &= 0 \\ 1 \leq x_1^2 &\leq 4 \\ \mu_2 &\geq 0 \end{aligned}$$

Poiché $-2\mu_2x_1 = 1$, $\mu_2 \neq 0$, e quindi $x_1^2 = 4$. D'altra parte, $-2\mu_2x_1 = 1$ e $\mu_2 > 0$ significa che $x_1 < 0$. Quindi l'unico punto candidato è $H = (-2, 0)$, che corrisponde a $\mu_H = (0, 1/4, 1)$.

P⁴ ($\mu_1 > 0$ e $g_1 = 0 \Rightarrow x_1^2 = 1$)

$$\begin{aligned} 2(\mu_1 - \mu_2)x_1 &= 1 \\ 1 - \mu_3 &= 0 \Rightarrow \mu_3 = 1 \\ x_1^2 &= 1 \\ \mu_2(x_1^2 - 4) &= 0 \Rightarrow \mu_2 = 0 \\ x_1^2 - 4 &\leq 0 \\ \mu_2 &\geq 0 \end{aligned}$$

Da $2(\mu_1 - \mu_2)x_1 = 1$ e $\mu_2 = 0$ deriva $2\mu_1x_1 = 1$, che impone $x_1 > 0$. Quindi l'unico punto candidato è $E = (1, 0)$, con $\mu_E = (1/2, 0, 1)$.

Fra i due punti candidati, è punto di ottimo globale H :

$$\begin{cases} E = (1, 0) \Rightarrow f(E) = 1 \\ H = (-2, 0) \Rightarrow f(H) = -2 \end{cases}$$

Esercizio 3

Risolvere con le condizioni di KKT il seguente problema:

$$\begin{aligned}\min f(x) &= x_1^2 + x_2^2 \\ g_1(x) &= x_1^2 + x_2^2 - 4 \leq 0 \\ g_2(x) &= -x_1 - x_2 - 2 \leq 0\end{aligned}$$

Soluzione

La Figura 4.10 rappresenta la regione delle soluzioni ammissibili.

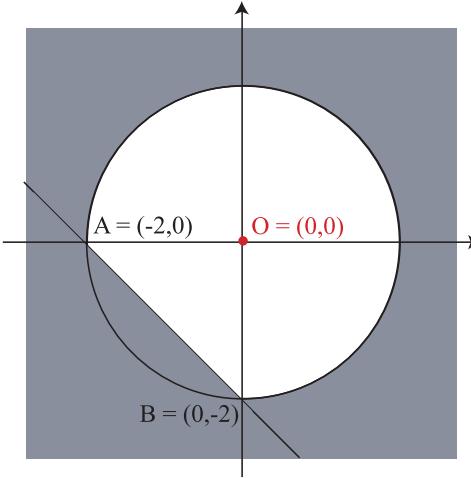


Figura 4.10: Regione ammissibile

Punti non regolari

I punti interni alla regione di ammissibilità sono tutti regolari. I punti per i quali passa un vincolo solo lo sono ovunque il gradiente è diverso da zero. I gradienti dei due vincoli sono:

$$\begin{aligned}\nabla g_1(x) &= [2x_1 \ 2x_2]' \\ \nabla g_2(x) &= [-1 \ -1]'\end{aligned}$$

Il primo si annulla solo nell'origine, dove il vincolo non è attivo; il secondo non si annulla mai. I due vincoli si intersecano in

$$\begin{aligned}g_1(x) &= x_1^2 + x_2^2 - 4 = 0 \\ g_2(x) &= -x_1 - x_2 - 2 = 0\end{aligned}$$

cioè in $A = (-2, 0)$ e $B = (0, -2)$. I loro gradienti sono

$$\nabla g_1(A) = [-4 \ 0]' \quad \nabla g_2 \equiv [-1 \ -1]' \quad \nabla g_1(B) = [0 \ -4]' \quad \nabla g_2 \equiv [-1 \ -1]'$$

linearmente indipendenti. Quindi tutti i punti sono regolari.

Condizioni di Karush-Kuhn-Tucker

La lagrangiana generalizzata è $\ell(x, \mu) = x_1^2 + x_2^2 + \mu_1(x_1^2 + x_2^2 - 4) + \mu_2(-x_1 - x_2 - 2)$. Quindi le condizioni necessarie del primo ordine sono:

$$\begin{aligned}\partial\ell/\partial x_1 &= 2x_1 + 2\mu_1 x_1 - \mu_2 = 0 \\ \partial\ell/\partial x_2 &= 2x_2 + 2\mu_1 x_2 - \mu_2 = 0 \\ \mu_1 g_1 &= \mu_1(x_1^2 + x_2^2 - 4) = 0 \\ \mu_2 g_2 &= \mu_2(-x_1 - x_2 - 2) = 0 \\ g_1 &= x_1^2 + x_2^2 - 4 \leq 0 \\ g_2 &= -x_1 - x_2 - 2 \leq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0\end{aligned}$$

Sottraendo i primi due vincoli, si ottiene

$$2(x_1 - x_2)(1 + \mu_1) = 0$$

Poiché $\mu_1 \geq 0 \Rightarrow 1 + \mu_1 \geq 1 > 0$, è $x_1 = x_2 = x$.

$$\begin{aligned}2x(1 + \mu_1) &= \mu_2 \\ \mu_1(x^2 - 2) &= 0 \\ \mu_2(x + 1) &= 0 \\ x^2 &\leq 2 \\ x &\geq -1 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0\end{aligned}$$

Il vincolo più semplice è $\mu_2(x + 1) = 0$, per cui spezziamo il problema originario P^0 in P^1 ($\mu_2 = 0$) e P^2 ($\mu_2 > 0$ e $x = -1$).

P¹ ($\mu_2 = 0$)

$$\begin{aligned}2x(1 + \mu_1) &= 0 \\ \mu_1(x^2 - 2) &= 0 \\ x^2 &\leq 2 \\ x &\geq -1 \\ \mu_1 &\geq 0\end{aligned}$$

Poiché $\mu_1 \geq 0$, è $x = 0$. Questo corrisponde al punto candidato $O = (0, 0)$, con $\mu_0 = (0, 0)$. Si noti che nessun vincolo è attivo in O , e corrispondentemente tutti i moltiplicatori sono nulli.

P² ($\mu_2 > 0$ e $x_1 = x_2 = x = -1$)

$$\begin{aligned}-2(1 + \mu_1) &= \mu_2 \\ \mu_1 &= 0 \\ x_1 = x_2 &= -1 \\ x^2 &\leq 2 \\ x &\geq -1 \\ \mu_1 &\geq 0\end{aligned}$$

che però implica $\mu_2 < 0$, e quindi non è ammissibile. In effetti, nel punto $(-1, -1)$ il gradiente della funzione obiettivo $\nabla f(-1, -1) = [-2 \ -2]'$ è parallelo al gradiente dell'unico vincolo attivo $\nabla g_2 \equiv [-1 \ -1]'$, ma è diretto nello stesso verso anziché in verso opposto. Questo corrisponde al segno errato di μ_2 .

Si ha quindi un solo punto candidato, che è il punto di ottimo globale: $O = (0, 0)$, con $f(O) = 0$.

Esercizio 4

Risolvere con le condizioni di Karush-Kuhn-Tucker il problema

$$\begin{aligned}\min f(x) &= x_1 + x_2 \\ h_1(x) &= x_1^2 - x_2 = 0 \\ g_1(x) &= x_1 \leq 0\end{aligned}$$

Soluzione

La Figura La Figura 4.11 rappresenta la regione delle soluzioni ammissibili.

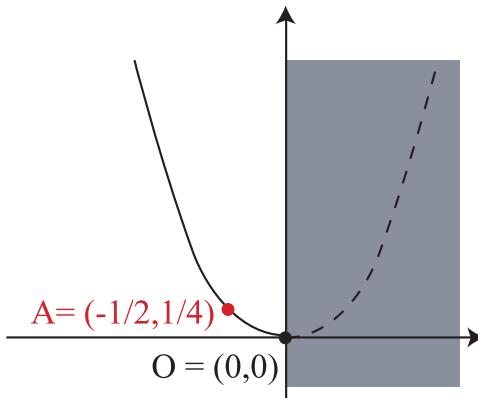


Figura 4.11: Regione ammissibile

Punti non regolari

I gradienti dei vincoli sono

$$\nabla h(x) = [2x_1 \ - 1]' \quad \nabla g(x) = [1 \ 0]'$$

e non si annullano mai. L'unico punto potenzialmente non regolare, quindi, è l'intersezione dei due vincoli, cioè l'origine, dove entrambi i vincoli sono attivi. In essa, però, i due vettori

$$\nabla h(O) = [0 \ - 1]' \quad \nabla g(O) = [1 \ 0]'$$

sono linearmente indipendenti. Quindi tutti i punti sono regolari.

Condizioni di Karush-Kuhn-Tucker

La lagrangiana generalizzata è $\ell(x, \lambda, \mu) = x_1 + x_2 + \lambda(x_1^2 - x_2) + \mu x_1$.

$$\begin{aligned}\partial \ell / \partial x_1 &= 1 + 2\lambda x_1 + \mu = 0 \\ \partial \ell / \partial x_2 &= 1 - \lambda = 0 \Rightarrow \lambda = 1 \\ \partial \ell / \partial \lambda = h &= x_1^2 - x_2 = 0 \\ \mu g &= \mu x_1 = 0 \\ g &= x_1 \leq 0\end{aligned}$$

da cui $\mu = -2x_1 - 1$

$$\begin{aligned}x_2 &= x_1^2 \\(2x_1 + 1)x_1 &= 0 \\x_1 &\leq 0\end{aligned}$$

da cui $x_1 = -1/2$ e $x_1 = 0$. La prima soluzione è accettabile e corrisponde al punto candidato $A = (-1/2, 1/4)$ con $\lambda = 1$ e $\mu = 0$. La seconda non è accettabile perché il punto $x = (0, 0)$ corrisponde ai moltiplicatori $\lambda = 1$ e $\mu = -1 < 0$, che non è ammissibile.

Quindi l'unico punto candidato è $A = (-1/2, 1/4)$, ed è il punto di ottimo globale.

Esercizio 5

Risolvere con le condizioni di Karush-Kuhn-Tucker il problema

$$\begin{aligned} \min f(x) &= x_2 \\ g_1(x) &= (x_1 - 1)^3 + (x_2 - 2) \leq 0 \\ g_2(x) &= (x_1 - 1)^3 - (x_2 - 2) \leq 0 \\ g_3(x) &= -x_1 \leq 0 \end{aligned}$$

Soluzione

La Figura 4.12 rappresenta la regione delle soluzioni ammissibili.

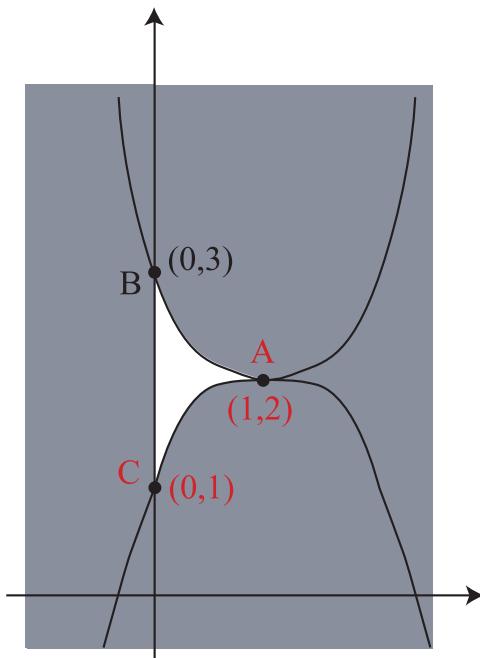


Figura 4.12: Regione ammissibile

Punti non regolari

I punti regolari sono quei punti nei quali i gradienti dei vincoli attivi sono fra loro linearmente indipendenti. Si noti che vanno considerati solo i vincoli attivi. Quindi, i punti interni alla regione di ammissibilità sono tutti regolari, quelli per i quali passa un vincolo solo lo sono ovunque il gradiente è diverso da zero, e coi via.

I gradienti dei tre vincoli sono:

$$\begin{aligned} \nabla g_1(x) &= \left[3(x_1 - 1)^2 \ 1 \right]' \\ \nabla g_2(x) &= \left[3(x_1 - 1)^2 \ -1 \right]' \\ \nabla g_3(x) &= [-1 \ 0]' \end{aligned}$$

Poiché nessun gradiente si annulla mai, ci interessano solo i punti in cui sono attivi almeno due vincoli. Vi sono tre punti di questo genere, indicati in figura con A , B e C .

$$\begin{aligned} g_1(A) = 0 &\Rightarrow (x_1 - 1)^3 = -(x_2 - 2) \\ g_2(A) = 0 &\Rightarrow (x_1 - 1)^3 = +(x_2 - 2) \end{aligned}$$

da cui $-(x_2 - 2) = (x_2 - 2)$, cioè $x_2 = 2$ e $(x_1 - 1)^3 = 0$, cioè $x_1 = 1$. In breve, $A = (1, 2)$.

$$\begin{aligned} g_1(B) = 0 &\Rightarrow (x_1 - 1)^3 = -(x_2 - 2) \\ g_3(B) = 0 &\Rightarrow x_1 = 0 \end{aligned}$$

da cui $-1 = 2 - x_2$, cioè $x_2 = 3$ e quindi $B = (0, 3)$. Infine,

$$\begin{aligned} g_2(C) = 0 &\Rightarrow (x_1 - 1)^3 = (x_2 - 2) \\ g_3(C) = 0 &\Rightarrow x_1 = 0 \end{aligned}$$

da cui $1 = 2 - x_2$, cioè $x_2 = 1$ e quindi $C = (0, 1)$.

In A è $\nabla g_1(A) = [0 \ 1]'$ e $\nabla g_2(A) = [0 \ -1]'$, cioè i due gradienti sono opposti. Dunque sono linearmente dipendenti: A non è regolare. Nei punti non regolari, le condizioni di Karush-Kuhn-Tucker perdono di validità, cioè non danno informazioni. Quindi, siamo costretti a considerare tutti i punti non regolari come punti candidati.

In B è $\nabla g_1(B) = [3 \ 1]'$ e $\nabla g_3(B) = [-1 \ 0]'$. Questi due vettori sono indipendenti. Anziché valutare il determinante della matrice composta accostando i due vettori, si può usare la definizione e cercare una combinazione lineare dei due vettori che si annulli. In altre parole, si impone $\alpha_1 \nabla g_1(B) + \alpha_2 \nabla g_3(B) = 0$ e si cercano i valori di α_1 e α_2 che soddisfano la condizione: si ottiene $[3\alpha_1 - \alpha_2 \ \alpha_1]'$, che è nullo solo per $\alpha_1 = \alpha_2 = 0$. Quindi i due vettori sono linearmente indipendenti.

Infine, in C è $\nabla g_2(C) = [3 \ -1]'$ e $\nabla g_3(C) = [-1 \ 0]'$. Questi due vettori sono indipendenti. Infatti imponendo $\alpha_1 \nabla g_2(C) + \alpha_2 \nabla g_3(C) = 0$ si ottiene $[3\alpha_1 - \alpha_2 \ -\alpha_1]'$, che è nullo solo per $\alpha_1 = \alpha_2 = 0$.

Condizioni di Karush-Kuhn-Tucker

La lagrangiana generalizzata è $\ell(x, \mu) = x_2 + \mu_1 (x_1 - 1)^3 + \mu_1 (x_2 - 2) + \mu_2 (x_1 - 1)^3 - \mu_2 (x_2 - 2) - \mu_3 x_1$, per cui

$$\begin{aligned} \partial \ell / \partial x_1 &= 3\mu_1 (x_1 - 1)^2 + 3\mu_2 (x_1 - 1)^2 - \mu_3 = 0 \\ \partial \ell / \partial x_2 &= 1 + \mu_1 - \mu_2 = 0 \\ \mu_1 g_1 &= \mu_1 \left[(x_1 - 1)^3 + (x_2 - 2) \right] = 0 \\ \mu_2 g_2 &= \mu_2 \left[(x_1 - 1)^3 - (x_2 - 2) \right] = 0 \\ \mu_3 g_3 &= -\mu_3 x_1 = 0 \\ g_1 \leq 0 &= (x_1 - 1)^3 + (x_2 - 2) \leq 0 \\ g_2 \leq 0 &= (x_1 - 1)^3 - (x_2 - 2) \leq 0 \\ g_3 \leq 0 &= -x_1 \leq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \\ \mu_3 &\geq 0 \end{aligned}$$

Conviene scegliere il vincolo più semplice, cioè il terzo e spezzare il problema originario P^0 in P^1 ($\mu_3 = 0$ e $g_3 \leq 0$) e P^2 ($\mu_3 > 0$ e $g_3 = 0$).

P¹ ($\mu_3 = 0$ e $g_3 \leq 0$)

$$\begin{aligned} 3\mu_1(x_1 - 1)^2 + 3\mu_2(x_1 - 1)^2 &= 0 \\ 1 + \mu_1 - \mu_2 &= 0 \\ \mu_1 \left[(x_1 - 1)^3 + (x_2 - 2) \right] &= 0 \\ \mu_2 \left[(x_1 - 1)^3 - (x_2 - 2) \right] &= 0 \\ (x_1 - 1)^3 + (x_2 - 2) &\leq 0 \\ (x_1 - 1)^3 - (x_2 - 2) &\leq 0 \\ x_1 &\geq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \end{aligned}$$

Il primo vincolo diventa $3(\mu_1 + \mu_2)(x_1 - 1)^2 = 0$. Si noti che μ_1 e μ_2 sono entrambi non negativi, per cui $(\mu_1 + \mu_2) \geq 0$ e ovviamente $(x_1 - 1)^2 \geq 0$: il loro prodotto si annulla solo quando si annulla uno dei due termini.

A questo punto potremmo dividere il problema P^1 in due sottoproblemi nel modo standard, considerando il prodotto $\mu_1 g_1 = 0$ oppure $\mu_2 g_2 = 0$. Tuttavia, l'osservazione che abbiamo appena fatto ci permette di fare qualcosa di meglio. Anziché seguire la via standard, ramifichiamo rispetto al vincolo $3(\mu_1 + \mu_2)(x_1 - 1)^2 = 0$, ottenendo i problemi P^3 ($\mu_1 + \mu_2 = 0$ e $(x_1 - 1)^2 \geq 0$, che è banale) e P^4 ($\mu_1 + \mu_2 > 0$ e $(x_1 - 1)^2 = 0$). Il vantaggio è che il problema P^3 si semplifica brutalmente, dato che μ_1 e μ_2 sono entrambi ≥ 0 , per cui la loro somma si può annullare solo se si annullano entrambi.

P³ ($\mu_1 + \mu_2 = 0$, cioè $\mu_1 = 0$ e $\mu_2 = 0$)

$$\begin{aligned} 1 &= 0 \\ x_1 &\geq 0 \end{aligned}$$

Il sottoproblema non ha soluzioni ammissibili.

P⁴ ($\mu_1 + \mu_2 > 0$ e $x_1 = 1$)

$$\begin{aligned} 1 + \mu_1 - \mu_2 &= 0 \\ \mu_1(x_2 - 2) &= 0 \\ -\mu_2(x_2 - 2) &= 0 \\ (x_2 - 2) &\leq 0 \\ -(x_2 - 2) &\leq 0 \\ x_1 &\geq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \end{aligned}$$

da cui $\mu_2 = \mu_1 + 1 \geq 1 > 0$. Quindi $-\mu_2(x_2 - 2) = 0$ implica $x_2 = 2$. Per μ_1 va bene qualsiasi valore non negativo. La soluzione $A = (1, 2)$ soddisfa le condizioni di Karush-Kuhn-Tucker, e quindi è punto candidato. A rigore, questo punto va escluso dalla trattazione, dato che non è regolare, e quindi è candidato in partenza.

P² ($\mu_3 > 0$ e $g_3 = 0$, cioè $x_1 = 0$)

$$\begin{aligned} 3\mu_1 + 3\mu_2 - \mu_3 &= 0 \\ 1 + \mu_1 - \mu_2 &= 0 \\ \mu_1 [-1 + (x_2 - 2)] &= 0 \\ \mu_2 [-1 - (x_2 - 2)] &= 0 \\ -1 + (x_2 - 2) &\leq 0 \\ -1 - (x_2 - 2) &\leq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \\ \mu_3 &> 0 \end{aligned}$$

Poiché $\mu_2 = 1 + \mu_1 \geq 1 > 0$, il quarto vincolo ($\mu_2(1 - x_2) = 0$) diviene $x_2 = 1$, da cui $\mu_1[-1 + (x_2 - 2)] = -2\mu_1 = 0$ per $\mu_1 = 0$, che implica $\mu_2 = 1$ e $\mu_3 = 3$. La soluzione $C = (0, 1)$ soddisfa le condizioni di Karush-Kuhn-Tucker, e quindi è punto candidato.

Fra i due punti candidati, quello di ottimo globale è C perché

$$\begin{cases} A = (1, 2) \Rightarrow f(A) = 2 \\ C = (0, 1) \Rightarrow f(C) = 1 \end{cases}$$

Esercizio 6

Risolvere con le condizioni di Karush-Kuhn-Tucker il problema

$$\begin{aligned}\min f(x) &= -x_1 \\ g_1(x) &= (x_1 - 1)^3 + (x_2 - 2) \leq 0 \\ g_2(x) &= (x_1 - 1)^3 - (x_2 - 2) \leq 0 \\ g_3(x) &= -x_1 \leq 0\end{aligned}$$

Soluzione

La Figura 4.13 rappresenta la regione delle soluzioni ammissibili.

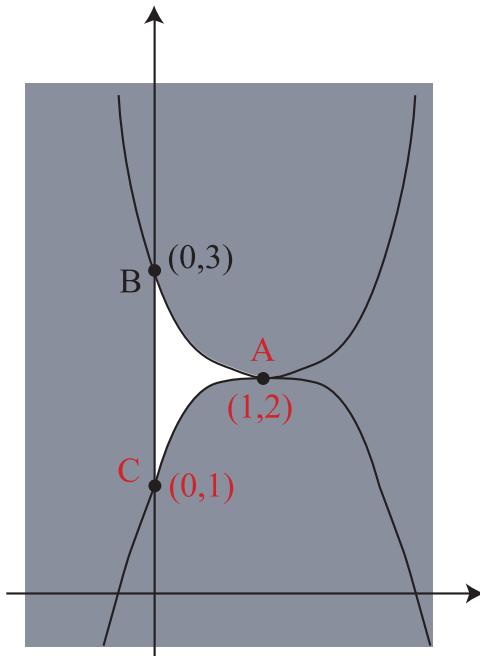


Figura 4.13: Regione ammissibile

Punti non regolari

Il problema ha gli stessi vincoli, e quindi la stessa regione ammissibile dell'Esercizio 6. Ripetiamo quindi l'analisi, e concludiamo che l'unico punto non regolare è $A = (1, 2)$.

Condizioni di Karush-Kuhn-Tucker

La lagrangiana generalizzata è $\ell(x, \mu) = -x_1 + \mu_1 (x_1 - 1)^3 + \mu_2 (x_2 - 2) + \mu_3 (x_1 - 1)^3 - \mu_2 (x_2 - 2) - \mu_3 x_1$, per cui

$$\begin{aligned}\partial\ell/\partial x_1 &= -1 + 3\mu_1 (x_1 - 1)^2 + 3\mu_2 (x_1 - 1)^2 - \mu_3 = 0 \\ \partial\ell/\partial x_2 &= \mu_1 - \mu_2 = 0 \\ \mu_1 g_1 &= \mu_1 [(x_1 - 1)^3 + (x_2 - 2)] = 0 \\ \mu_2 g_2 &= \mu_2 [(x_1 - 1)^3 - (x_2 - 2)] = 0 \\ \mu_3 g_3 &= -\mu_3 x_1 = 0 \\ g_1 \leq 0 &= (x_1 - 1)^3 + (x_2 - 2) \leq 0 \\ g_2 \leq 0 &= (x_1 - 1)^3 - (x_2 - 2) \leq 0 \\ g_3 \leq 0 &= -x_1 \leq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \\ \mu_3 &\geq 0\end{aligned}$$

Poiché $\mu_1 = \mu_2 = \mu$, il primo vincolo diventa $6\mu(x_1 - 1)^2 = \mu_3 + 1 \geq 1 > 0$, da cui deriva che $\mu > 0$. Sommando e sottraendo il terzo e il quarto vincolo, si ottiene che $x_1 = 1$ e $x_2 = 2$. Quindi, il primo vincolo richiederebbe $\mu_3 = -1$ e il quinto vincolo richiederebbe $\mu_3 = 0$, un assurdo. Ne segue che nessun punto soddisfa le condizioni di Karush-Kuhn-Tucker.

D'altra parte, un minimo globale esiste senz'altro, perché la regione ammissibile è chiusa e limitata e la funzione obiettivo continua. Il punto di ottimo globale è A , che non essendo regolare può essere ottimo anche senza soddisfare le condizioni del primo ordine. Si noti come invece nell'esercizio precedente lo stesso punto, pur essendo non regolare, soddisfacesse le condizioni: entrambi i casi possono presentarsi. Per questo, occorre comunque tenere in considerazione i punti non regolari come candidati.

Esercizio 7

Risolvere con le condizioni di Karush-Kuhn-Tucker il problema

$$\begin{aligned} \min z &= (x_1 + 1)^2 + \left(x_2 + \frac{1}{2}\right)^2 \\ x_1^2 - x_2^2 &\leq 0 \\ x_1 - x_2 &\leq 0 \end{aligned}$$

Soluzione

Questo problema ha una regione ammissibile molto particolare, che consiste nel quadrante superiore compreso fra le bisettrici degli assi, ma anche nella semiretta $x_2 = x_1$ con $x_1 \leq 0$. La Figura 4.5 rappresenta la regione delle soluzioni ammissibili.

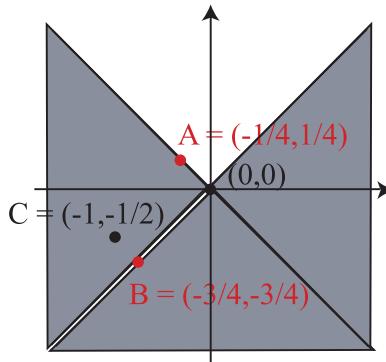


Figura 4.14: Regione ammissibile

Punti non regolari

I gradienti dei vincoli $g_1(x) = x_1^2 - x_2^2 \leq 0$ e $g_2(x) = x_1 - x_2 \leq 0$ sono

$$\nabla g_1 = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix} \quad \nabla g_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

I punti in cui nessuno dei due vincoli è attivo sono regolari per definizione. Quelli in cui è attivo solo g_1 , cioè quelli della bisettrice del secondo e quarto quadrante, esclusa l'origine ($x_2 = -x_1$, con $x_1 \neq 0$) sono regolari fin tanto che il gradiente non si annulla, ma questo richiedebbe $x_1 = x_2 = 0$, che è escluso. Quindi tutti questi punti sono regolari. In nessun punto è attivo solo g_2 . I punti in cui sono attivi entrambi i vincoli, cioè quelli della bisettrice del primo e terzo quadrante ($x_2 = x_1 = \xi$) sono tutti non regolari, perché in essi $\nabla g_1 = [2\xi - 2\xi]^T$ e $\nabla g_2 = [1 - 1]^T$.

Condizioni di Karush-Kuhn-Tucker

Le condizioni analitiche del primo ordine si ottengono a partire dalla lagrangiana $\ell(x, \mu) = (x_1 + 1)^2 + (x_2 + \frac{1}{2})^2 + \mu_1(x_1^2 - x_2^2) + \mu_2(x_1 - x_2)$.

$$\begin{aligned}\frac{\partial \ell}{\partial x_1} &= 2(x_1 + 1) + 2\mu_1 x_1 + \mu_2 = 0 \\ \frac{\partial \ell}{\partial x_2} &= 2\left(x_2 + \frac{1}{2}\right) - 2\mu_1 x_2 - \mu_2 = 0 \\ \mu_1 g_1(x) &= \mu_1(x_1^2 - x_2^2) = 0 \\ \mu_2 g_2(x) &= -\mu_2(x_1 - x_2) = 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \\ g_1(x) &\leq 0 \\ g_2(x) &\leq 0\end{aligned}$$

Dato che sono punti non regolari, e quindi automaticamente candidati, escludiamo a priori dalla trattazione i punti della bisettrice del primo e terzo quadrante: $x_1 - x_2 < 0$. Ne consegue, dal quarto vincolo, che $\mu_2 = 0$: il moltiplicatore di un vincolo non attivo (g_2) è sempre nullo.

$$\begin{aligned}2(x_1 + 1) + 2\mu_1 x_1 &= 0 \\ 2\left(x_2 + \frac{1}{2}\right) - 2\mu_1 x_2 &= 0 \\ \mu_1(x_1 + x_2) &= 0 \\ 0 &= 0 \\ \mu_1 &\geq 0 \\ \mu_2 &= 0 \\ x_1 + x_2 &\geq 0 \\ x_1 - x_2 &< 0\end{aligned}$$

Dividiamo il problema in due sottoproblemi in base al terzo vincolo.

Problema $\mu_1 = 0$ Dai primi due vincoli, si ottiene il punto $C = (-1, -1/2)$. Si tratta di una soluzione non ammissibile per il vincolo $g_1 \leq 0$, per cui va scartata. D'altra parte, si tratta di una soluzione sensata, dato che è il punto di minimo della funzione obiettivo in assenza di vincoli.

Problema $\mu_1 > 0$ Risulta $x_2 = -x_1$ e quindi

$$\begin{aligned}2(x_1 + 1) + 2\mu_1 x_1 &= 0 \\ 2\left(-x_1 + \frac{1}{2}\right) + 2\mu_1 x_1 &= 0\end{aligned}$$

che sottraendo i due vincoli dà $x_1 = -1/4$ e $x_2 = 1/4$, mentre $\mu_1 = 3$. Questo è l'unico punto candidato suggerito dalle condizioni di Karush-Kuhn-Tucker.

Si tratta ora di decidere fra il punto $A = (-1/4, 1/4)$ e i punti della retta $x_2 = x_1 = \alpha$. Determiniamo il punto di minimo fra questi e lo confrontiamo con A .

$$\min f(\alpha) = (\alpha + 1)^2 + \left(\alpha + \frac{1}{2}\right)^2$$

La soluzione si ottiene annullando la derivata rispetto al parametro α : $f'(\alpha) = 2(\alpha + 1) + 2\left(\alpha + \frac{1}{2}\right) = 0$, da cui $\alpha = -3/4$ e quindi $B = (-3/4, -3/4)$, dove la funzione obiettivo vale $f(B) = 1/8$. In A , invece, la funzione obiettivo vale $f(A) = 9/8$. Quindi l'ottimo globale sta nel punto non regolare $B = (-3/4, -3/4)$.

Parte II

Modelli con preferenze complesse

Capitolo 5

La preferenza paretiana

Questo capitolo tratta una situazione decisionale più complessa: continuiamo ad assumere che vi sia un solo decisore e un solo scenario, ma la relazione di preferenza non sarà più un ordine debole descritto da una funzione valore (o costo) conforme.

Per poter dire qualcosa di utile, dobbiamo fare ipotesi sulla struttura della relazione di preferenza, e ovviamente ipotesi diverse produrranno modelli diversi. Storicamente, il primo modello, che è ancora uno dei più importanti, fu quello proposto nella seconda metà dell'Ottocento da Pareto¹. Ricordiamo che abbiamo supposto che l'impatto abbia valori vettoriali a p dimensioni ($F \subseteq \mathbb{R}^p$).

Definizione 14 Si dice preferenza paretiana la relazione:

$$\Pi = \{(f, f') : f_l \leq f'_l \text{ per ogni } l \in \{1, \dots, p\}\}$$

ovvero

$$f \preceq f' \Leftrightarrow f_l \leq f'_l \text{ per ogni } l \in \{1, \dots, p\}$$

Nella teoria di Pareto, un impatto è preferibile a un altro quando tutti gli elementi del primo non superano i corrispondenti elementi del secondo. Sostanzialmente, si assume che gli indicatori f_i rappresentino dei costi. Ovviamente, esiste anche un caso paretiano nel quale essi rappresentano utilità e la relazione di preferenza viene definita corrispondentemente; qui adottiamo una convenzione uniforme con quella adottata per trattare la Programmazione Matematica.

Il caso paretiano è un modello abbastanza ragionevole di situazioni pratiche: se gli indicatori rappresentano costi, è normale che si preferiscano impatti con indicatori bassi a impatti con indicatori alti. Tuttavia, in questo modello si possono confrontare solo impatti con elementi che si dominano tutti nello stesso verso. Si parla di *ipotesi degli indicatori incomparabili*, intendendo che il decisore non è in grado di paragonare fra loro due impatti quando indicatori diversi danno suggerimenti opposti. Questa limitazione può rendere il modello paretiano di preferenza irrealistico.

5.1 Proprietà formali della preferenza paretiana

Teorema 13 La relazione di preferenza paretiana è un ordine parziale.

¹Vilfredo Pareto (1848-1923), economista, matematico e ingegnere italiano, nato in Francia e vissuto in Italia e Svizzera.

Dimostrazione. Infatti è

- riflessiva: $f \preceq f$ per ogni $f \in F$

$$f_l = f_l \text{ per ogni } l \in \{1, \dots, p\} \Rightarrow f_l \leq f_l \text{ per ogni } l \in \{1, \dots, p\} \Rightarrow f \preceq f$$

- transitiva: $f \preceq f'$ e $f' \preceq f'' \Rightarrow f \preceq f''$ per ogni $f, f', f'' \in F$

$$\begin{cases} f \preceq f' \\ f' \preceq f'' \end{cases} \Rightarrow \begin{cases} f_l \leq f'_l \text{ per ogni } l \in \{1, \dots, p\} \\ f'_l \leq f''_l \text{ per ogni } l \in \{1, \dots, p\} \end{cases} \Rightarrow \\ \Rightarrow f_l \leq f''_l \text{ per ogni } l \in \{1, \dots, p\} \Rightarrow f \preceq f''$$

- antisimmetrica: $f \preceq f'$ e $f' \preceq f \Rightarrow f = f'$ per ogni $f, f' \in F$

$$\begin{cases} f \preceq f' \\ f' \preceq f \end{cases} \Rightarrow \begin{cases} f_l \leq f'_l \text{ per ogni } l \in \{1, \dots, p\} \\ f'_l \leq f_l \text{ per ogni } l \in \{1, \dots, p\} \end{cases} \Rightarrow \\ \Rightarrow f_l = f'_l \text{ per ogni } l \in \{1, \dots, p\} \Rightarrow f = f'$$

■ Le dimostrazioni sono semplici, perché sfruttano le proprietà dei numeri reali, di cui godono i singoli indicatori presi uno per uno. Sembrano banali giochi di notazione, ma non lo sono, perché il passaggio dalla relazione fra impatti (\preceq) alla relazione fra indicatori (\leq) è cruciale, ed è consentito solo dalla definizione di Pareto.

Osservazione 2 In generale, la relazione di preferenza paretiana non è completa.

Esempio 17 Un banale controesempio alla completezza è fornito dai due vettori

$$f = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad f' = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

dove f è meglio di f' per il primo indicatore, ma è peggio per il secondo. Quindi la preferenza paretiana non vale in alcuno dei due versi: i due impatti sono incomparabili ($f \bowtie f'$).

5.2 Dominanza paretiana

Come discusso nella Sezione 3.6, la relazione di preferenza stretta fra impatti induce una relazione di dominanza fra soluzioni, nel senso che una soluzione ne domina un'altra se e solo se l'impatto della prima è strettamente preferibile all'impatto della seconda. Si usa la preferenza stretta perché in generale possono esistere diverse soluzioni con lo stesso impatto, e quindi indifferenti fra loro; se si dicesse che tali soluzioni si dominano a vicenda, bisognerebbe scartarle tutte, e questo non pare ragionevole.

Nel caso paretiano, la dominanza fra soluzioni si esprime dunque come segue:

$$x' \prec x \Leftrightarrow f(x') \prec f(x) \Leftrightarrow \begin{cases} f_l(x') \leq f_l(x) \text{ per ogni } l \in \{1, \dots, p\} \\ \exists \bar{l} \in \{1, \dots, p\} : f_{\bar{l}}(x') < f_{\bar{l}}(x) \end{cases}$$

e questo consente di dividere le soluzioni ammissibili $x \in X$ in due gruppi, uno dei quali andrà scartato in qualsiasi decisione ragionevole, mentre l'altro andrà preso in considerazione.

Definizione 15 Si dice soluzione dominata *qualsiasi soluzione ammissibile* $x \in X$ che ammetta un'altra soluzione $x' \in X$ che la domina:

$$\exists x' \in X, \bar{l} \in \{1, \dots, p\} : \begin{cases} f_l(x') \leq f_l(x) & \text{per ogni } l \in \{1, \dots, p\} \\ f_{\bar{l}}(x') < f_{\bar{l}}(x) \end{cases}$$

Definizione 16 Si dice soluzione paretiana *qualsiasi soluzione ammissibile* $x \in X$ tale che nessun'altra soluzione x' la domina:

$$\forall x' \in X, \exists \bar{l} \in \{1, \dots, p\} : f_{\bar{l}}(x) < f_{\bar{l}}(x') \text{ oppure } f(x') = f(x)$$

Definizione 17 Si dice regione paretiana X° l'insieme delle soluzioni paretiane.

La regione paretiana è il corrispondente dell'insieme dei punti di ottimo globale in Programmazione Matematica, con due differenze piccole, ma non trascurabili, dovute alla incompletezza della relazione di preferenza paretiana:

1. le soluzioni paretiane non sono preferibili a tutte le altre, mentre i punti di ottimo globale lo sono;
2. le soluzioni paretiane non sono tutte indifferenti fra loro, mentre i punti di ottimo globale lo sono.

Questo significa che, mentre in Programmazione Matematica di solito basta trovare un punto di ottimo globale e si possono ignorare gli altri, nel caso paretiano è opportuno trovare tutte le soluzioni paretiane.

5.3 Determinazione della regione paretiana

Il problema fondamentale a questo punto risulta la determinazione della regione paretiana. Esistono diversi modi per ottenerla, che però hanno ambiti applicativi diversi, vantaggi e svantaggi diversi e producono risultati diversi. Questo perché, invece di generare esattamente la regione paretiana X° , in generale ne forniscono una sottostima (cioè un sottoinsieme) o una sovrastima (cioè un sovrainsieme). Non esiste un metodo che valga in generale e produca esattamente X° . Nel seguito descriviamo:

1. l'applicazione diretta della *definizione*, che vale solo per il caso finito;
2. il *metodo della trasformazione inversa*, che vale solo se gli indicatori sono due ($f \in \mathbb{R}^2$);
3. le *condizioni di Karush-Kuhn-Tucker*, che valgono solo per il caso continuo e producono una sovrastima di X° ;
4. il *metodo dei pesi*, che vale sempre, ma produce una sottostima di X° ;
5. il *metodo dei vincoli*, che vale sempre, ma produce una sovrastima di X° .

5.3.1 L'applicazione della definizione

Nel caso finito, la regione paretiana si può trovare in tempo finito, applicando la definizione stessa, ovvero costruendo il grafo della dominanza fra soluzioni, dove i nodi sono le soluzioni e gli archi la relazione di dominanza, come si è fatto nella Sezione 3.6 trattando relazioni di preferenza generiche. In questo grafo, le soluzioni paretiane sono i nodi privi di archi entranti.

Esempio 18 Riprendiamo in esame il problema del viaggio introdotto nell’Esempio 5 del Capitolo 3 e supponiamo che l’impatto f sia descritto completamente dal tempo di viaggio f_1 (misurato in ore) e dal costo del viaggio f_2 (misurato in Euro), con i valori riportati in Tabella 5.1. Gli indicatori sono entrambi intesi come funzioni di costo e sono fra loro incomparabili.

	Tempo (ore)	Costo (Euro)
Treno	5.5	100
Auto	4.0	150
Aereo	1.0	300
Pullman	5.0	180
Taxi	4.0	400

Tabella 5.1: Tempi e costi associati a diversi mezzi di trasporto per compiere un viaggio

La relazione di preferenza che ne deriva è

$$\Pi = \{(Auto, Pullman), (Auto, Taxi), (Aereo, Taxi)\}$$

Le soluzioni paretiane sono quindi Treno, Aereo e Auto. Ricordiamo che nell’Esempio 5 della Sezione 3.6 si era assunto di saper confrontare Auto e Treno, cosicché quest’ultimo risultava dominato. La differenza è che qui la relazione di preferenza è costruita meccanicamente in base ai valori dei due indicatori, mentre lì era data elencando esplicitamente ciascuna coppia di impatti.

Esempio 19 Consideriamo la seguente matrice di valutazione, nella quale gli indicatori sono tutti da massimizzare, anziché da minimizzare.

Alternative	Indicatori		
	f_1	f_2	f_3
a_1	100	60	60
a_2	70	75	25
a_3	60	40	20
a_4	40	100	80
a_5	20	80	100

Procedendo per confronti a coppie, si verifica che l’alternativa a_3 è dominata dalle alternative a_1 e a_2 . Le altre quattro alternative sono tutte paretiane.

5.4 Metodo della trasformazione inversa

Questo metodo consiste anch’esso nell’applicare la definizione, ma è applicabile anche a problemi continui, purché rappresentabili graficamente nello spazio degli indicatori. Esso infatti consiste in una risoluzione grafica del problema, e richiede di:

- determinare l’inversa $\phi : F \rightarrow X$ della trasformazione $f : X \rightarrow F$;
- sostituire $x = \phi(f)$ nell’espressione dei vincoli in modo da ottenere un’espressione analitica della regione F nello spazio degli indicatori;
- determinare il sottoinsieme F° degli impatti non dominati, cioè dell’immagine della regione paretiana X° nello spazio degli indicatori; questi impatti sono quelli il cui quadrante inferiore sinistro non interseca F ;

- determinare attraverso la trasformazione inversa $\phi : F \rightarrow X$ la regione paretiana X° nello spazio delle variabili.

Dato che in genere si deve determinare F° graficamente, il metodo è applicabile solo nel caso in cui vi siano due indicatori ($p = 2$).

Esempio 20 Si consideri il seguente problema:

$$\begin{aligned} \min f_1(x) &= x_1 + x_2 \\ \min f_2(x) &= -x_1 \\ g_1(x) &= 3x_1^2 + 4x_2 - 12 \leq 0 \\ g_2(x) &= -x_2 \leq 0 \end{aligned}$$

La Figura 5.1 illustra la regione ammissibile e gli antigradianti dei due indicatori.

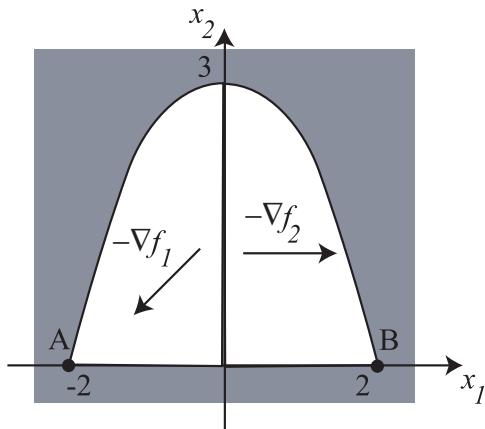


Figura 5.1: La regione ammissibile del problema

Anzi tutto, determiniamo la trasformazione inversa $\phi : F \rightarrow X$:

$$\begin{cases} f_1(x) = x_1 + x_2 \\ f_2(x) = -x_1 \end{cases} \Rightarrow \begin{cases} x_1(f) = -f_2 \\ x_2(f) = f_1 - x_1 = f_1 + f_2 \end{cases}$$

da cui risulta la regione F (rappresentata in Figura 5.2):

$$\begin{cases} g_1(x(f)) = 3f_2^2 + 4(f_1 + f_2) - 12 \leq 0 \\ g_2(x(f)) = -f_1 - f_2 \leq 0 \end{cases} \Rightarrow \begin{cases} f_1 \leq -\frac{3}{4}f_2^2 - f_2 + 3 \\ f_1 + f_2 \geq 0 \end{cases}$$

È facile individuare graficamente l'immagine della regione paretiana: basta trovare i punti che non ammettono altri punti nel quadrante in basso a sinistra. Tali punti costituiscono il segmento CD , con $C = (-2, 2)$ e $D = (2, -2)$.

Ora dobbiamo applicare la trasformazione inversa e tornare nello spazio delle variabili. Poiché la trasformazione inversa è lineare, è possibile fare una discussione semplificata: il segmento CD nello spazio degli indicatori corrisponde a un altro segmento nello spazio delle variabili di decisione. Per trovare questo segmento, basta trasformare i due estremi: la regione paretiana X° è il segmento compreso fra $x(C) = A = (-2, 0)$ e $x(D) = B = (2, 0)$.

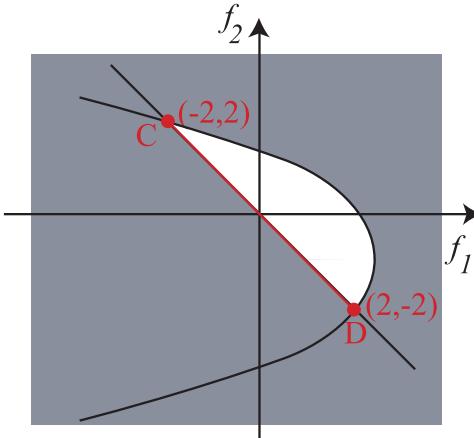


Figura 5.2: L'insieme degli impatti del problema (in rosso l'immagine della regione paretiana)

In generale, dovremmo invece esprimere la regione F° in forma parametrica come $f(\alpha)$ e sottoporla alla trasformazione inversa in modo da ottenere l'espressione parametrica $x(f(\alpha)) = x(\alpha)$. In questo caso

$$\begin{cases} f_1(\alpha) = \alpha \\ f_2(\alpha) = -\alpha \end{cases} \Rightarrow \begin{cases} x_1(\alpha) = f_2(\alpha) = \alpha \\ x_2(\alpha) = -f_1(\alpha) - f_2(\alpha) = -\alpha + \alpha = 0 \end{cases}$$

con $\alpha \in [-2; 2]$.

5.5 Le condizioni di Karush-Kuhn-Tucker

Queste condizioni estendono le analoghe condizioni definite per la Programmazione Matematica. Sono ancora una volta condizioni locali, per cui passano attraverso una definizione di dominanza paretiana locale, anziché globale. E sono ancora condizioni necessarie, ma non sufficienti, e quindi producono un insieme X^{KKT} di soluzioni candidate ad essere paretiane, ma non certamente tali. Tale insieme sarà quindi in generale più ampio della regione paretiana: $X^{\text{KKT}} \supseteq X^\circ$. Bisognerà poi filtrare via in una fase successiva i punti dominati.

Nel seguito, descriviamo schematicamente il procedimento, che segue le tracce di quello già descritto nel Capitolo 4. Per prima cosa, si introduce una descrizione analitica della regione ammissibile X attraverso vincoli e ipotesi di regolarità sulle funzioni che definiscono impatto e vincoli:

$$X = \{x \in \mathbb{R}^n : g_j(x) \leq 0 \text{ per } j \in \{1, \dots, m\}\} \subseteq \mathbb{R}^n$$

e

$$f = [f_1 \ f_2 \ \dots \ f_p]^T$$

con $g_j(x) \in C^1(X)$ funzioni regolari per $j \in \{1, \dots, m\}$ e $f_l \in C^1(X)$ per $l \in \{1, \dots, p\}$.

Soluzioni localmente paretiane

Definizione 18 Si dice soluzione localmente paretiana una soluzione ammisible $x \in X$ che ammette un intorno $\mathcal{U}_{x,\epsilon}$ entro il quale nessun'altra soluzione la domina:

$$\exists \mathcal{U}_{x,\epsilon} : x' \not\prec x \text{ per ogni } x' \in \mathcal{U}_{x,\epsilon} \cap X$$

Una soluzione paretiana è localmente paretiana, ma non vale l'inverso, come i punti di ottimo globale sono punti di ottimo locale, ma non viceversa.

La Figura 5.3 rappresenta in grigio l'insieme degli impatti F di un problema decisionale nello spazio degli indicatori \mathbb{R}^p . In questo spazio è facile vedere se un impatto f corrisponde a un punto x dominato o paretiano: basta costruire il quadrante inferiore sinistro a partire da f (cioè l'insieme dei punti con coordinate $\leq f_l$ per ogni $l \in \{1, \dots, p\}$) e verificare che tale quadrante non contenga punti di F diversi da f . Per esempio, in Figura 5.3 le soluzioni A e B sono paretiane. Le soluzioni C e D sono dominate, dato che qualsiasi intorno degli impatti $f(C)$ e $f(D)$ contiene punti con coordinate non superiori e almeno una coordinata strettamente inferiore (per esempio quelli del segmento compreso fra $f(B)$ e $f(C)$ sono migliori di $f(C)$ rispetto all'indicatore f_1 e ugualmente buoni rispetto a f_2 ²). Infine, il punto E è localmente paretiano, dato che nessun impatto nell'intersezione fra lo spazio degli impatti F e l'immagine dell'intorno $\mathcal{U}_{E,\epsilon}$ raffigurata dal cerchio cade nel quadrante inferiore sinistro di $f(E)$, delimitato dalle linee tratteggiate. Tuttavia, E non è un punto globalmente paretiano, perché l'impatto $f(E)$ è dominato da parecchi altri impatti, fra cui $f(A)$, $f(B)$, $f(C)$ e $f(D)$.

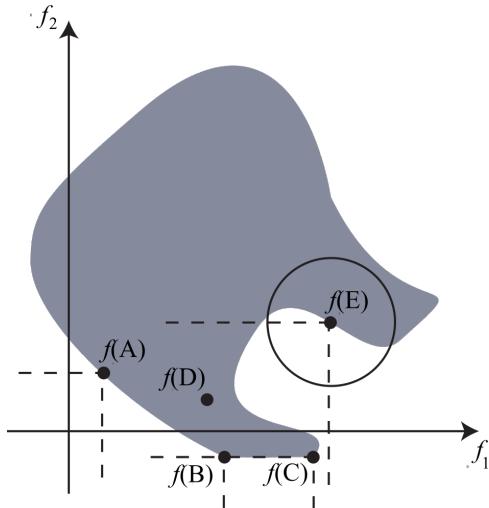


Figura 5.3: A e B sono globalmente paretiani; C è dominato (più precisamente, è debolmente paretiano); D è dominato; E è localmente, ma non globalmente paretano.

Nel caso in cui X è discreto (vedi Figura 5.4) tutte le soluzioni ammissibili sono punti isolati. Questo significa che, assegnando valori abbastanza piccoli a ϵ , l'intorno $\mathcal{U}_{x,\epsilon}$ intersecca X nel solo punto x . Quindi, in un problema discreto ogni soluzione è localmente paretiana. Lo stesso avveniva in Programmazione Matematica. Questo è coerente con il fatto che in generale le condizioni di KKT restituiscono una sovrastima dell'insieme delle soluzioni paretiane, ma nel caso discreto esse restituiscono tutto X e sono quindi completamente inutili.

²Il punto C si definisce debolmente paretiano perché non è possibile migliorare tutti gli indicatori restando nell'insieme F , ma ai fini del corso questo concetto non ci interessa.

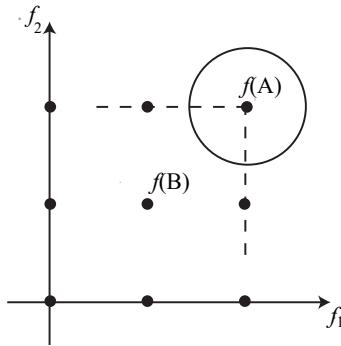


Figura 5.4: Nei problemi discreti, tutti i punti ammissibili sono localmente paretiani

Traccia della dimostrazione

Lo sviluppo delle condizioni è analogo a quello del caso a singolo obiettivo.

1. Definizione Se un punto x è localmente paretiano, tutti gli archi ξ ammissibili in x sono non miglioranti.

2. Condizione di ammissibilità Gli archi ammissibili vengono sostituiti dalle relative direzioni tangenti p_ξ , che nei punti regolari sono tutte e sole le direzioni con $(\nabla g_j(x))^T p \leq 0$ per tutti i vincoli attivi ($g_j(x) = 0$). Anche in questo caso, consideriamo tutti i punti non regolari come candidati, perché in essi le condizioni sono inaffidabili.

3. Condizione di non miglioramento Ora sostituiamo tutti gli indicatori con le loro approssimazioni del primo ordine attraverso la forma di Taylor. Le direzioni nelle quali l'indicatore f_l non migliora rispettano la condizione $(\nabla f_l(x))^T p \geq 0$. Come in Programmazione Matematica, ammettiamo l'uguaglianza perché i termini di ordine superiore potrebbero far peggiorare l'indicatore anche lungo direzioni ortogonali al gradiente. Queste condizioni sono solo necessarie e possono classificare come candidati anche punti dominati (vedremo un esempio più avanti).

4. Prima interpretazione geometrica Combinando ammissibilità e non miglioramento, si ha che, se \tilde{x} è un punto regolare e localmente paretiano e p una direzione tale che $(\nabla g_j)^T p \leq 0$ per tutti i vincoli attivi in \tilde{x} (cioè un vettore tangente ad arco ammissibile), allora esiste un indicatore f_l tale che $(\nabla f_l)^T p \geq 0$ (cioè un indicatore che migliora strettamente lungo l'arco).

Geometricamente, le direzioni p con prodotti scalari $(\nabla g_j)^T p \leq 0$ per tutti i vincoli attivi formano il cono ammissibile (vedi Sezione 4.2.3), mentre le direzioni con $(\nabla f_l)^T p \leq 0$ per ogni indicatore formano anch'esse un cono, composto dalle direzioni in cui tutti gli indicatori non peggiorano. Nel caso di un solo indicatore, tali direzioni formavano un semispazio, che abbiamo chiamato semispazio migliorante. Per estensione, chiameremo *cono migliorante* il nuovo insieme. Il teorema afferma che nessuna direzione del primo cade all'interno del secondo: *il cono ammissibile e il cono migliorante non hanno punti interni comuni*; possono al massimo toccarsi, perché nelle direzioni limite, con prodotti scalari nulli, le condizioni non sono decisive.

Esempio 21 Si consideri il seguente problema:

$$\begin{aligned} \min f_1 &= -2x_1 - x_2 \\ \min f_2 &= -x_1 - 2x_2 \\ g_1(x) &= -x_1 \leq 0 \\ g_2(x) &= -x_2 \leq 0 \\ g_3(x) &= x_1^2 + x_2 - 4 \leq 0 \end{aligned}$$

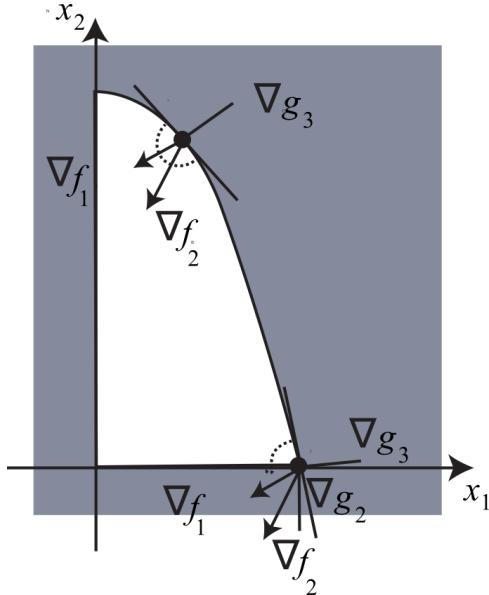


Figura 5.5: Interpretazione geometrica delle condizioni di KKT: il cono ammissibile e il cono migliorante non devono avere direzioni interne comuni

Calcoliamo i gradienti degli indicatori e dei vincoli:

$$\begin{aligned} \nabla f_1(x) &= \begin{bmatrix} -2 \\ -1 \end{bmatrix} & \nabla f_2(x) &= \begin{bmatrix} -1 \\ -2 \end{bmatrix} \\ \nabla g_1(x) &= \begin{bmatrix} -1 \\ 0 \end{bmatrix} & \nabla g_2(x) &= \begin{bmatrix} 0 \\ -1 \end{bmatrix} & \nabla g_3(x) &= \begin{bmatrix} 2x_1 \\ 1 \end{bmatrix} \end{aligned}$$

La Figura 5.5 mostra due situazioni tipiche³: nel punto $A = (2, 0)$ il cono ammissibile interseca il cono migliorante. Infatti, spostandosi da A verso l'alto a sinistra si trova un cono di direzioni che permettono di rimanere in X abbassando il valore sia di f_1 sia di f_2 . Nel punto $B = (1/2, 15/4)$, invece, i due coni non si intersecano. Infatti, spostandosi da B in qualsiasi direzione o si esce da X o si peggiora almeno uno dei due obiettivi.

5. Lemma di Farkas L'applicazione del lemma di Farkas è un po' più complicata che per la Programmazione Dinamica, dato che non si ha più un vettore da una parte e un cono dall'altra, ma due coni. Il trucco è introdurre un vettore ausiliario e applicare il lemma due volte: la prima volta al vettore ausiliario e al cono ammissibile, la seconda all'opposto del vettore ausiliario e al cono migliorante.

³LA FIGURA VA COMPLETATA E CORRETTA

L'esistenza di due coni che non si intersecano, ma al massimo si toccano, equivale all'esistenza di un iperpiano che li separa. Se γ è uno dei due vettori normali a questo iperpiano, uno dei due coni sta dalla stessa parte di γ rispetto all'iperpiano e l'altro dalla parte opposta. Senza perdita di generalità, supponiamo che il cono ammissibile stia dalla parte opposta a γ , cioè che

$$p^T \gamma \leq 0 \text{ per ogni } p : p^T \nabla g_j(x) \leq 0$$

Il lemma di Farkas impone che γ stia nel cono dei gradienti dei vincoli attivi:

$$\exists \mu_j \geq 0 : \gamma = \sum_{j=1}^m \mu_j \nabla g_j(x) \text{ per ogni } j : g_j(x) = 0$$

Da parte sua, il vettore $-\gamma$ sta dalla parte opposta ai gradienti degli indicatori:

$$p^T (-\gamma) \leq 0 \text{ per ogni } p : p^T \nabla f_l(x) \leq 0 \text{ per ogni } l$$

e quindi, per il lemma di Farkas, $-\gamma$ sta nel cono dei gradienti degli indicatori:

$$\exists w_l \geq 0 : -\gamma = \sum_{l=1}^p w_l \nabla f_l$$

Sommendo le due equazioni, si ottiene la tesi:

$$\exists w_l \geq 0, \mu_j \geq 0 : \sum_{l=1}^p w_l \nabla f_l + \sum_j \mu_j \nabla g_j = 0$$

Queste sono le condizioni di KKT per il caso paretiano.

6. Seconda interpretazione geometrica Geometricamente, le condizioni di KKT per la Programmazione Matematica, si interpretavano dicendo che l'antigradiente della funzione obiettivo deve cadere nel cono dei gradienti dei vincoli attivi. Nel caso paretiano, le condizioni di KKT si possono interpretare dicendo che *il cono degli antigradianti degli indicatori interseca il cono dei gradienti dei vincoli attivi*.

Esempio 22 Tornando alla Figura 5.5, si può osservare come in A i due coni non si intersechino, mentre in B lo fanno. Infatti:⁴

7. Forma standard delle condizioni di KKT Come in Programmazione Matematica, si possono reintrodurre i vincoli non attivi nel sistema imponendo che i loro moltiplicatori siano nulli, grazie alle solite condizioni di complementarietà $\mu_j g_j(x) = 0$, e si possono trattare i vincoli di uguaglianza senza raddoppiarli esplicitamente attraverso l'uso di moltiplicatori liberi in segno.

Si noti poi che, dato un punto x e moltiplicatori w_l e μ_j che soddisfano le condizioni, lo stesso punto x soddisfa le condizioni con i moltiplicatori αw_l e $\alpha \mu_j$ per ogni fattore $\alpha > 0$. In altre parole, i moltiplicatori sono determinati a meno di una costante moltiplicativa. Si puo allora impostare una condizione di normalizzazione. Generalmente, si immagina di dividere tutti i moltiplicatori per la somma $\sum_{l=1}^p w_l$, in modo che

$$\sum_{l=1}^p w_l = 1$$

Ovviamente, questo richiede di dimostrare che $\sum_l w_l > 0$, ma questo è vero, perché nei punti regolari i gradienti dei vincoli attivi sono linearmente indipendenti, e quindi $\sum_j \mu_j \nabla g_j \neq 0$.

⁴L'ESEMPIO VA COMPLETATO: I COEFFICIENTI μ e w CHE CONFERMANO QUESTA INTERSEZIONE SONO (per esempio) $\mu_3 = 1$ e $w_1 = w_2 = 1/3$.

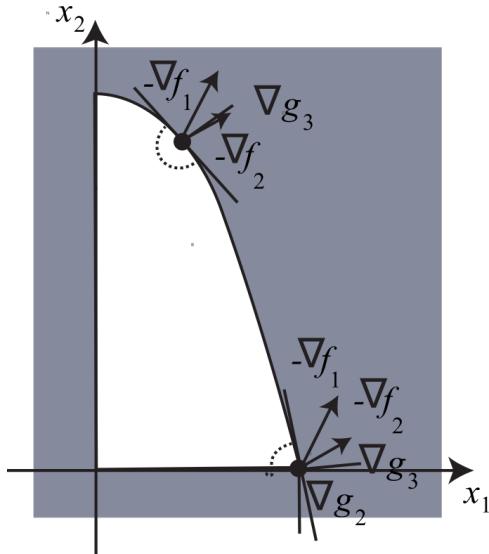


Figura 5.6: Seconda interpretazione geometrica delle condizioni di KKT: il cono degli antigradienti degli indicatori deve intersecare il cono dei gradienti dei vincoli attivi

Esempio 23 APPLICAZIONE DI QUANTO SOPRA AI PUNTI A E B DELL'ESEMPIO

La regione paretiana come ipersuperficie parametrica

In Programmazione Matematica, applicare le condizioni di KKT significa determinare n variabili di decisione x_i e m moltiplicatori μ_j , risolvendo un sistema di $n+m$ equazioni: le n equazioni che annullano le derivate parziali della lagrangiana e le m equazioni che annullano i prodotti $\mu_j g_j$. Avendo lo stesso numero di equazioni e incognite⁵, il sistema ha in generale un numero di soluzioni finito. Le $2m$ disequazioni che impongono l'ammissibilità dei punti candidati e la non negatività dei moltiplicatori μ_j possono incidere riducendo tale numero.

Nel caso paretiano, si tratta di determinare n variabili di decisione x_i , m moltiplicatori μ_j e p moltiplicatori w_l , risolvendo un sistema di $n+m+1$ equazioni (l'ultima è la condizione di normalizzazione sui w_l). In generale, si otterranno quindi ∞^{p-1} soluzioni, che possiamo interpretare come una ipersuperficie parametrica a $p-1$ dimensioni espressa nella forma $x = x(w)$.

Esempio 24 Si consideri il seguente problema:

$$\begin{aligned} \min f_1 &= -2x_1 - x_2 \\ \min f_2 &= -x_1 - 2x_2 \\ g_1(x) &= -x_1 \leq 0 \\ g_2(x) &= -x_2 \leq 0 \\ g_3(x) &= x_1^2 + x_2 - 4 \leq 0 \end{aligned}$$

⁵Il numero rimane lo stesso se i vincoli di uguaglianza vengono gestiti attraverso moltiplicatori liberi in segno, poiché in tal caso ci sono s variabili λ aggiuntive e s vincoli di uguaglianza $h_i(x) = 0$

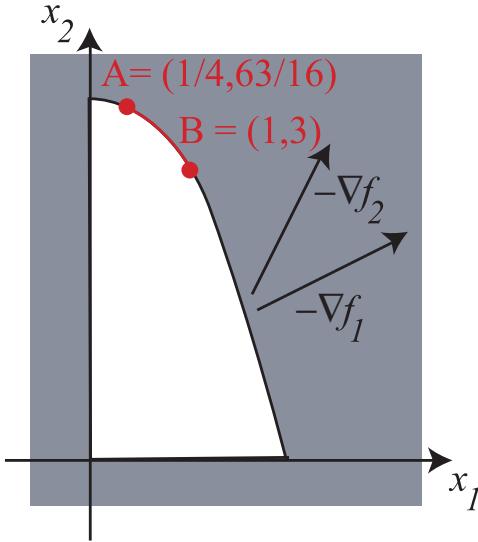


Figura 5.7: Ricerca della regione paretiana

La Figura 5.7 ne rappresenta la regione ammissibile. I due indicatori f_1 e f_2 sono costi, da minimizzare, e i loro gradienti sono

$$\nabla f_1(x) = \begin{bmatrix} -2 \\ -1 \end{bmatrix} \quad \nabla f_2(x) = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

mentre le frecce nella figura indicano gli antigradienti, cioè le direzioni di miglioramento. I gradienti dei vincoli sono:

$$\nabla g_1(x) = \begin{bmatrix} -1 \\ -0 \end{bmatrix} \quad \nabla g_2(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \nabla g_3(x) = \begin{bmatrix} 2x_1 \\ 1 \end{bmatrix}$$

Poiché $w_1 + w_2 = 1$, possiamo porre $w_1 = w$ e $w_2 = 1 - w$. Le diseguaglianze $w_1 \geq 0$ e $w_2 \geq 0$ diventano quindi $0 \leq w \leq 1$. Il sistema delle condizioni di KKT si riduce quindi a:

$$\begin{aligned} -2w - 1 + w - \mu_1 + 2\mu_3 x_1 &= 0 \\ -w - 2 + 2w - \mu_2 + \mu_3 &= 0 \\ \mu_1 g_1(x) = -\mu_1 x_1 &= 0 \\ \mu_2 g_2(x) = -\mu_2 x_2 &= 0 \\ \mu_3 g_3(x) = -\mu_3 (x_1^2 + x_2 - 4) &= 0 \\ 0 \leq w \leq 1 & \\ \mu_1 \geq 0 & \\ \mu_2 \geq 0 & \\ \mu_3 \geq 0 & \\ g_1(x) = -x_1 \leq 0 & \\ g_2(x) = -x_2 \leq 0 & \\ g_3(x) = x_1^2 + x_2 - 4 \leq 0 & \end{aligned}$$

Dalla prima condizione, $2\mu_3 x_1 = w + 1 + \mu_1 \geq 1 > 0$. Quindi $x_1 > 0$ e $\mu_3 > 0$. La terza condizione garantisce allora che $\mu_1 = 0$ e la quinta che $x_2 = 4 - x_1^2$. Ora sappiamo che la regione paretiana è un arco della parabola $g_3(x) = 0$.

$$\begin{aligned}
x_1 &= \frac{w+1}{2\mu_3} \\
\mu_2 &= w - 2 + \mu_3 \\
\mu_1 &= 0 \\
\mu_2(4 - x_1^2) &= 0 \\
0 &\leq w \leq 1 \\
\mu_2 &\geq 0 \\
\mu_3 &> 0 \\
0 &< x_1 \leq 2
\end{aligned}$$

Scomponiamo il problema P_0 nei sottoproblemi P_1 ($\mu_2 = 0$) e P_2 ($\mu_2 > 0$ e $4 - x_1^2 = 0$)

P_1 ($\mu_2 = 0$) Dal secondo vincolo, si ricava $\mu_3 = 2-w$, e quindi la linea parametrica in (x_1, x_2) (ipersuperficie a $p-1 = 1$ dimensioni):

$$\begin{cases} x_1 = \frac{w+1}{2(2-w)} \\ x_2 = 4 - \left[\frac{w+1}{2(2-w)} \right]^2 \end{cases} \quad \text{con } w \in [0; 1]$$

Si tratta di un arco della parabola $x_2 = 4 - x_1^2$. Per sapere di quale arco si tratti, dobbiamo determinare quali punti corrispondono ai valori di w compresi fra 0 e 1. Per $w = 0$, è $x(0) = A = (1/4, 63/16)$. Per $w = 1$, è $x(1) = B = (1, 3)$. Per essere sicuri che valori intermedi di w corrispondano a punti intermedi, basta osservare che $x_1(w)$ è monotona crescente, dato che

$$\frac{dx_1}{dw} = \frac{1 \cdot 2(2-w) + 2(w+1)}{4(2-w)^2} = \frac{3}{2(2-w)^2} > 0$$

P_2 ($\mu_2 > 0$ e $4 - x_1^2 = 0 \Rightarrow x_1 = 2$)

$$\begin{aligned}
\mu_3 &= \frac{w+1}{4} \\
\mu_2 &= w - 2 + \frac{w+1}{4} = \frac{5w-7}{4}
\end{aligned}$$

Poiché $w \in [0; 1]$, $\mu_2 < 0$, che non è ammissibile. Quindi, questo problema non determina punti candidati ad essere localmente paretiani.

Dal punto di vista geometrico, si può notare che, in tutti i punti della regione paretiana, il cono delle direzioni ammissibili non interseca il cono delle direzioni miglioranti. Inoltre, il cono dei gradienti dei vincoli interseca il cono degli antigradieni degli indicatori.

Nell'esempio seguente, le condizioni di KKT determinano un insieme strettamente più ampio della regione paretiana. Questo conferma che si tratta solo di condizioni necessarie.

Esempio 25 Si consideri il seguente problema:

$$\begin{aligned}\min f_1 &= -x_2 \\ \min f_2 &= -x_1 - x_2 \\ g_1(x) &= x_1 - 1 \leq 0 \\ g_2(x) &= -x_1 \leq 0 \\ g_3(x) &= -x_1^2 + x_2 - 1 \leq 0\end{aligned}$$

La Figura 5.8 rappresenta la regione ammissibile del problema. Le frecce indicano gli antigradienti degli indicatori, cioè le direzioni di più rapido miglioramento.

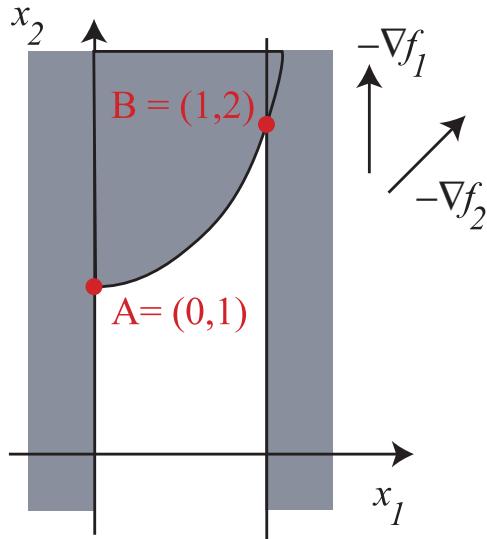


Figura 5.8: Ricerca della regione paretiana: il punto A rispetta le condizioni di KKT, e quindi è candidato, pur non essendo nemmeno localmente paretiano

I gradienti delle funzioni obiettivo sono:

$$\nabla f_1(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \nabla f_2(x) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

mentre i gradienti dei vincoli sono

$$\nabla g_1(x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \nabla g_2(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \nabla g_3(x) = \begin{bmatrix} -2x_1 \\ 1 \end{bmatrix}$$

Come nell'esempio precedente, teniamo conto in partenza della normalizzazione

dei moltiplicatori w_l . Il sistema complessivo da risolvere è:

$$\begin{aligned} w - 1 + \mu_1 - \mu_2 - 2\mu_3 x_1 &= 0 \\ -w + w - 1 + \mu_3 &= 0 \\ \mu_1 g_1(x) = \mu_1 (x_1 - 1) &= 0 \\ \mu_2 g_2(x) = -\mu_2 x_1 &= 0 \\ \mu_3 g_3(x) = -\mu_3 (-x_1^2 + x_2 - 1) &= 0 \\ 0 \leq w \leq 1 & \\ \mu_1 \geq 0 & \\ \mu_2 \geq 0 & \\ \mu_3 \geq 0 & \\ g_1(x) = x_1 - 1 \leq 0 & \\ g_2(x) = -x_1 \leq 0 & \\ g_3(x) = -x_1^2 + x_2 - 1 \leq 0 & \end{aligned}$$

Dalla seconda condizione, si ricava subito $\mu_3 = 1$, per cui il vincolo $g_3(x) = 0$ è attivo: la regione paretiana giace sulla parabola.

$$\begin{aligned} w - 1 + \mu_1 - \mu_2 - 2x_1 &= 0 \\ \mu_1 (x_1 - 1) &= 0 \\ \mu_2 x_1 &= 0 \\ x_2 &= x_1^2 + 1 \\ 0 \leq w \leq 1 & \\ \mu_1 \geq 0 & \\ \mu_2 \geq 0 & \\ 0 \leq x_1 \leq 1 & \end{aligned}$$

Scomponiamo il problema P_0 . Vista la forma delle condizioni, questa volta scomponiamo P_0 in tre sottoproblemi: P_1 ($x_1 = 0$ e $\mu_1 = 0$), P_2 ($0 < x_1 < 1$ e $\mu_1 = \mu_2 = 0$) e P_3 ($x_1 = 1$ e $\mu_2 = 0$).

P_1 ($x_1 = 0$ e $\mu_1 = 0$) Dal primo vincolo si ricava $\mu_2 = w - 1$. Poiché deve essere $\mu_2 \geq 0$, l'unico valore lecito risulta $w = 1$, da cui $\mu_2 = 0$. Il punto $A = (0, 1)$ è candidato, con $\mu = (0, 0, 1)$ e $w = 1$. Eppure, A non è nemmeno localmente paretiano, perché spostandosi verso destra si conserva f_1 costante e si migliora f_2 (e salendo leggermente si può persino migliorare f_1). In effetti, questo è coerente con le condizioni di KKT, dato che esse chiedono soltanto che almeno uno degli indicatori non migliori in base alla forma di Taylor arrestata al primo ordine. Per sicurezza, dobbiamo considerare candidato un punto in cui anche solo un indicatore rimane costante al primo ordine e tutti gli altri migliori.

P_2 ($0 < x_1 < 1$ e $\mu_2 = 0$) Dal primo vincolo, $\mu_1 = \mu_2 = 0$ implica $x_1 = (w - 1)/2$. Tuttavia, dovendo essere $0 < x_1 < 1$, occorrerebbe $1 < w < 3$, che non è ammissibile. Quindi, questo sottoproblema non fornisce punti candidati.

P_3 ($x_1 = 1$ e $\mu_2 = 0$) Dal primo vincolo si ricava $\mu_1 = 3 - w$, dal quarto $x_2 = 2$. Quindi, il punto $B = (1, 2)$ è candidato, con $\mu = (3 - w, 0, 1)$ per ogni $w \in [0; 1]$. Si tratta di un caso degenero, in cui infinite soluzioni delle condizioni di KKT corrispondono a un solo punto candidato.

In conclusione, il problema ha due punti candidati, di cui solo B è paretiano.

5.6 Metodo dei pesi

Il metodo dei pesi consiste nel combinare linearmente fra loro gli indicatori e minimizzare la combinazione stessa. Il teorema seguente dimostra che, per combinazioni con coefficienti strettamente positivi, si ottiene sempre una soluzione paretiana. È un metodo molto semplice e molto usato, ma vedremo che presenta diversi svantaggi.

Teorema 14 *Sia $z(x) = \sum_{l=1}^p w_l f_l(x)$ una combinazione convessa degli indicatori con coefficienti $w_l > 0$ per $l \in \{1, \dots, p\}$ e $\sum_{l=1}^p w_l = 1$. Se x° è punto di ottimo globale in X rispetto a $z(x)$, allora x° è punto globalmente paretiano rispetto all'impatto $f(x)$.*

Dimostrazione. Il teorema si dimostra per assurdo. Se x° non fosse un punto paretiano, per la Definizione 15 dovrebbe esistere un altro punto $x' \in X$ tale che $x' \prec x$, cioè $f_l(x') \leq f_l(x^\circ)$ per ogni $l \in \{1, \dots, p\}$ e un indice \bar{l} tale che $f_{\bar{l}}(x') < f_{\bar{l}}(x^\circ)$. Ma allora, sarebbe

$$\begin{cases} w_l f_l(x') \leq w_l f_l(x^\circ) & \text{per ogni } l \in \{1, \dots, p\} \setminus \{\bar{l}\} \\ w_{\bar{l}} f_{\bar{l}}(x') < w_{\bar{l}} f_{\bar{l}}(x^\circ) \end{cases}$$

Si noti che la seconda proprietà vale solo per $w_{\bar{l}}$ strettamente positivo. Sommando le p diseguaglianze, si ottiene

$$\sum_{l=1}^p w_l f_l(x') < \sum_{l=1}^p w_l f_l(x^\circ) \Rightarrow z(x') < z(x^\circ)$$

e quindi x° non può essere punto di ottimo globale per $z(x)$. Dalla contraddizione deriva la tesi. ■

Questo risultato pare molto simile alle condizioni di KKT, ma è invece molto diverso:

1. l'ipotesi del teorema richiede alla soluzione x° di essere un punto di ottimo globale, e non semplicemente di soddisfare le condizioni di KKT;
2. i coefficienti w_l usati nella combinazione devono essere strettamente positivi, e non semplicemente non negativi.

Per questi due motivi, le condizioni del metodo dei pesi sono molto più strette delle condizioni di KKT. Questo le rende sufficienti a dimostrare che un punto che le soddisfa è paretiano, anziché solo necessarie. Di conseguenza, il sottoinsieme X^{pesi} che ne risulta è un sottoinsieme di X° , anziché un sovrainsieme.

Osservazione 3 *Se nel Teorema 14 ci si limitasse a imporre che x° sia punto di ottimo globale, senza richiedere anche che i pesi w_l siano strettamente positivi, potrebbe esistere un punto x' con valori $f_l(x') = f_l(x^\circ)$ per gli indicatori di peso positivo e con valori $f_l(x') < f_l(x^\circ)$ per gli indicatori di peso nullo; il punto x° sarebbe ancora di ottimo globale per $z(x)$, ma sarebbe anche dominato da x' , e quindi la tesi non varrebbe più.*

Esempio 26 Si consideri ancora il problema dell'Esempio 24, rappresentato nella Figura 5.7. Applichiamo il metodo dei pesi combinando i due indicatori f_1 e f_2 con coefficienti $w_l > 0$. La funzione obiettivo

$$z(x) = w_1 f_1 + w_2 f_2$$

ha un antigradiente che è pari alla combinazione degli antigradienti ∇f_1 e ∇f_2 con gli stessi coefficienti w_1 e w_2 .

Al variare dei coefficienti, tale vettore cade sempre all'interno del cono individuato dai due antigradienti combinati. Questo suggerisce come risolvere per via grafica il problema: si ottiene un solo punto di ottimo globale, che al variare dei coefficienti scorre lungo l'arco AB della parabola nella Figura 5.7. Il metodo non fornisce tutta la regione paretiana, che sarebbe l'intero arco, ma solo l'arco privato dei due estremi. Infatti, usando coefficienti strettamente positivi non si ottengono i due punti estremi, che corrispondono a $w = [1 \ 0]^T$ e $w = [0 \ 1]^T$.

Esempio 27 Si consideri il problema descritto nell'Esempio 25. Applicando il metodo dei pesi, si ottiene per qualsiasi valore dei coefficienti w_l il punto B. Questo è corretto, perché tale punto è l'unico paretiano. Quindi in questo caso il sottoinsieme determinato dal metodo coincide con l'intera regione paretiana.

Per discutere i punti di forza e i limiti del metodo dei pesi, ci occorrono alcune definizioni.

Definizione 19 Data una soluzione paretiana $x^\circ \in X^\circ$, definiamo supporto di x° l'insieme $\text{Supp}(x^\circ) \subset \mathbb{R}^{p-1}$ dei vettori w tali che $w_l > 0$ per ogni $l \in \{1, \dots, p\}$, $\sum_{l=1}^p w_l = 1$ e x° è punto di ottimo globale per il problema $\min_{x \in X} \sum_{l=1}^p w_l f_l(x)$.

Definizione 20 Definiamo soluzione supportata una soluzione paretiana $x^\circ \in X^\circ$ il cui supporto sia non vuoto: $\text{Supp}(x^\circ) \neq \emptyset$.

Esempio 28 Consideriamo il problema dell'albero ricoprente minimo sul grafo completo K_3 (che è banalmente un triangolo con tre vertici e tre lati) con le due funzioni di costo indicate nella Tabella 5.2.

f_1	1	2	3	f_2	1	2	3
1	-	1	3	1	-	13	10
2	1	-	6	2	13	-	8
3	3	6	-	3	10	8	-

Tabella 5.2: Due funzioni di costo definite sui lati del grafo K_3

Le condizioni di KKT non sono utili in questo caso, perché il problema è discreto, e quindi tutte le soluzioni sono localmente paretiane. Il metodo della trasformazione inversa è in genere inapplicabile nei problemi combinatorici, perché X consiste in un numero esponenziale di soluzioni isolate nello spazio delle variabili di decisione, e corrisponde a un insieme esponenziale di impatti isolati nello spazio degli indicatori \mathbb{R}^2 . Trovare l'immagine della regione paretiana in questo spazio significa passare in rassegna tutti gli impatti possibili. Nel caso specifico, però, questo compito è facile, perché il problema ha solo $n = 3$ variabili di decisione, associate ai tre lati del grafo: $x_{12}, x_{13}, x_{23} \in \{0, 1\}$, e la regione ammissibile contiene solo tre soluzioni: $X = \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$. A queste tre soluzioni corrispondono tre impatti distinti, rappresentati nella Figura 5.9. I tre impatti hanno tutti quadrante inferiore

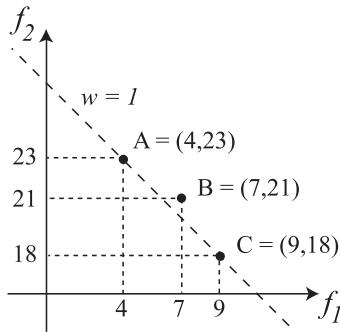


Figura 5.9: Le tre soluzioni ammissibili sono tutte globalmente paretiane, ma B , cioè l'albero $\{(1, 2), (2, 3)\}$ non è ottima per nessun valore di w

sinistro vuoto, e quindi le tre soluzioni ammissibili sono tutte globalmente paretiane.

Applichiamo ora il metodo dei pesi, combinando gli obiettivi f_1 e f_2 con i pesi w e $(1 - w)$, dove $w \in (0, 1)$, e risolvendo il problema dell'albero ricoprente minimo per ciascun valore di w . I costi dei lati sono

- $f(1, 2) = w \cdot 1 + (1 - w) \cdot 13 = 13 - 12w$
- $f(1, 3) = w \cdot 3 + (1 - w) \cdot 10 = 10 - 7w$
- $f(2, 3) = w \cdot 6 + (1 - w) \cdot 8 = 8 - 2w$

Cerchiamo la soluzione ottima con l'algoritmo di Prim, partendo dal vertice 1 e applicandolo in modo parametrico, cioè tenendo conto di tutti i valori possibili del parametro $w \in (0, 1)$ e suddividendo via via i casi e sottocasi che questo comporta⁶. Il lato di costo minimo incidente nel vertice 1 può essere $(1, 2)$ o $(1, 3)$ a seconda del valore di w , e in particolare del segno della differenza $f(1, 2) - f(1, 3) = (13 - 12w) - (10 - 7w) = 3 - 5w$:

1. se $w \leq 3/5$, costa meno il lato $(1, 3)$, che quindi diventa il primo lato dell'albero ricoprente minimo. A questo punto, dobbiamo trovare il lato di costo minimo nel taglio che separa i vertici 1 e 3 dal vertice 2, cioè il minimo fra $f(1, 2)$ e $f(3, 2)$:

$$f(1, 2) - f(3, 2) = (13 - 12w) - (8 - 2w) = 5 - 10w$$

Il confronto dà luogo a due sottocasi:

- (a) se $w \leq 1/2$, costa meno il lato $(3, 2)$, per cui l'albero ricoprente ottimo è $\{(1, 3), (3, 2)\}$ e costa $18 - 9w$
- (b) se $1/2 \leq w \leq 3/5$, costa meno il lato $(1, 2)$, per cui l'albero ricoprente ottimo è $\{(1, 3), (1, 2)\}$ e costa $23 - 19w$

2. se $w \geq 3/5$, costa meno il lato $(1, 2)$, che quindi diventa il primo lato dell'albero ricoprente minimo. A questo punto, dobbiamo trovare il lato di costo minimo nel taglio che separa 1 e 2 da 3, cioè il minimo fra $f(1, 3) = 10 - 7w$ e $f(2, 3) = 8 - 2w$.

$$f(1, 3) - f(2, 3) = (10 - 7w) - (8 - 2w) = 2 - 5w$$

Siccome questa espressione è sempre negativa per $w \geq 3/5$, l'albero ricoprente minimo è $\{(1, 2), (1, 3)\}$ e costa $23 - 19w$.

⁶Ovviamente, si potrebbe applicare qualsiasi algoritmo valido per il problema dell'albero ricoprente minimo, ad esempio l'algoritmo di Kruskal.

Riassumendo, si sono trovati due punti paretiani.

Tuttavia la combinazione convessa dei due obiettivi dà luogo per la terza soluzione a una funzione obiettivo combinata pari a $21 - 14w$, che è sempre più alta del valore delle altre due soluzioni paretiane. Quindi la terza soluzione non può essere determinata col metodo dei pesi: è non supportata.

5.6.1 Vantaggi e svantaggi

Il metodo dei pesi ha diversi vantaggi:

1. è applicabile a qualsiasi problema, anche quelli dove le condizioni di KKT sono inutili (ad es., i problemi discreti);
2. è molto intuitivo, dato che spesso i decisori trovano naturale combinare gli indicatori sommandoli, dopo averli espressi in unità di misura per le quali i loro valori numerici non siano troppo diversi;
3. richiede solo di costruire un obiettivo ausiliario per il problema, conservando ogni altro aspetto, fra cui in particolare la regione ammissibile; molto spesso, un algoritmo valido per ottimizzare i singoli indicatori rimane valido per ottimizzare l'obiettivo ausiliario.

Al tempo stesso, però, presenta diversi svantaggi:

1. richiede di considerare tutti i valori dei pesi w , i quali costituiscono un'infinità continua; questo comporta una risoluzione parametrica, che in genere non è banale, e per problemi di grandi dimensioni può diventare intrattabile;
2. per ciascun w richiede di trovare tutte le soluzioni globalmente ottime (non ne basta una sola come in Programmazione Matematica);
3. non fornisce l'intera regione paretiana X° , ma solo le soluzioni supportate, che ne possono costituire una sottostima anche molto ridotta.

Teorema 15 *Nei problemi combinatorici, il numero di soluzioni non supportate può crescere esponenzialmente con il numero n delle variabili decisionali; in corrispondenza, il metodo dei pesi può fornire una frazione della regione paretiana via via convergente a zero.*

Esempio 29 * Si può costruire abbastanza facilmente un esempio del Teorema 15 nel caso del problema del cammino minimo a due obiettivi. Applicando idee simili, si possono costruire esempi per altri problemi combinatorici con un numero qualsiasi di indicatori. Si consideri un grafo⁷ costituito da $2k + 3$ nodi: un nodo di origine s , un nodo intermedio i_0 , un nodo di destinazione t e k coppie di nodi i_r e i'_r con $r = 1, \dots, k$. Dal nodo s escono tre archi, diretti ai nodi i_k (di costo $f(s, i_k) = [2^k + 1 \ 0]^T$), i'_k (di costo $f(s, i'_k) = [0 \ 2^k + 2]^T$) e i_0 (di costo $f(s, i_0) = [1 \ 1]^T$). Da ciascuno dei nodi i_r con $r = 0, \dots, k-1$ e da ciascuno dei nodi i'_r con $r = 1, \dots, k-1$ escono due archi, diretti ai nodi della coppia successiva, i_{r+1} e i'_{r+1} , di costo $f(i_r, i_{r+1}) = f(i'_r, i_{r+1}) = [2^r \ 0]^T$ e $f(i_r, i'_{r+1}) = f(i'_r, i'_{r+1}) = [0 \ 2^r]^T$. Infine, dai nodi i_k e i'_k esce un solo arco, diretto al nodo t , di costo $f(i_k, t) = f(i'_k, t) = [2 \ 2]^T$. Per costruzione, esistono $2^k + 2$ cammini distinti dal nodo s al nodo t : il cammino (s, i_k, t) di costo $[2^k + 2 \ 0]^T$, il cammino (s, i'_k, t) di costo

*This example provides advanced concepts, that are not part of the course's syllabus.

⁷Che andrà disegnato in una futura versione delle dispense.

$[0 \ 2^k + 2]^T$ e 2^k cammini che vanno da s a i_0 e attraversano la sequenza delle coppie (i_r, i'_r) visitando uno e uno solo dei nodi di ciascuna coppia. Per costruzione, ciascuno di questi cammini ha costo $[h + 2 \ 2^k + 1 - h]^T$, con $h = 0, \dots, 2^k - 1$, cioè in questi cammini f_1 assume tutti i valori interi da 2 a $2^k + 1$, mentre f_2 corrispondentemente assume il valore complementare in modo che la loro somma sia $f_1 + f_2 = 2^k + 1$. Se si tracciano questi impatti sul piano $f_1 f_2$, ci si rende conto che la famiglia dei cammini che passano per i_0 ha come immagine i punti interi di un segmento della retta $f_1 + f_2 = 2^k + 1$, mentre gli altri due cammini rappresentano punti estremi che giacciono sugli assi f_1 e f_2 . Tutte le soluzioni sono paretiane. D'altra parte, per qualsiasi combinazione convessa di pesi, la soluzione ottima del problema a obiettivo singolo è uno dei due cammini estremi; le altre soluzioni sono non supportate. Il loro numero cresce esponenzialmente con k , cioè con il numero delle variabili decisionali del problema, che è pari al numero degli archi, $4k + 5$. La frazione delle soluzioni non supportate contiene 2^k soluzioni su $2^k + 2$, e ovviamente tende alla totalità delle soluzioni, al crescere del numero di variabili.

Per evitare la necessità di un algoritmo parametrico, si può pensare di *campionare* l'insieme dei pesi $w \in (0; 1)^p$ ed elencare le soluzioni ottime trovate per ciascun peso. Questo semplifica la procedura, perché si usa lo stesso algoritmo che ottimizza i singoli indicatori, ma introduce due nuove difficoltà:

1. non è nemmeno più garantito che si trovino tutte le soluzioni supportate, per cui si perdono ulteriori soluzioni paretiane;
2. è possibile che molti diversi valori campionati dei pesi producano la stessa soluzione paretiana, per cui il metodo può essere computazionalmente inefficiente rispetto a quello parametrico.

5.6.2 Il metodo dei pesi nella Programmazione Lineare*

Poiché i problemi di Programmazione Lineare hanno ottimo globale nei vertici della regione ammissibile (oppure sulle facce comprese fra vertici ottimi) e poiché le soluzioni paretiane fornite dal metodo dei pesi sono soluzioni ottime per qualche combinazione degli obiettivi con pesi w_l , tali soluzioni cadranno sempre sui vertici o tutt'al più sulle facce. Si può dimostrare che in questo caso il metodo fornisce l'intera regione paretiana, e che quindi ci si può limitare a determinare i vertici paretiani. A questo scopo, esistono metodi *ad hoc*, che non approfondiamo. Vediamo un esempio condotto con un metodo parametrico esplicito (analisi di casi e sottocasi).

Esempio 30 . Consideriamo il seguente problema con indicatori e vincoli lineari

$$\begin{aligned} \min f_1 &= -x_1 + 3x_2 \\ \min f_2 &= 4x_1 - x_2 \\ -x_1 + x_2 &\leq 7/2 \\ x_1 + x_2 &\leq 11/2 \\ 2x_1 + x_2 &\leq 9 \\ x_1 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

per il quale la Figura 5.10 riporta la regione ammissibile e le direzioni di miglioramento delle funzioni obiettivo.

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

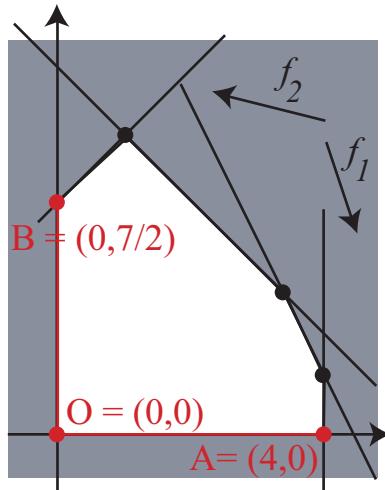


Figura 5.10: Nei problemi lineari, la regione paretiana coincide con quella parte della frontiera della regione ammissibile che è delimitata dai vertici che risultano ottimi globali per opportuni valori dei pesi

Il metodo dei pesi combina gli obiettivi con un vettore di coefficienti $w \in (0; 1)^p$, ottenendo una funzione obiettivo $z(x) = (4 - 5w)x_1 + (4w - 1)x_2$. Quindi, si risolve all'ottimo il problema risultante con l'algoritmo del simplex applicato in modo parametrico. Fortunatamente, siamo già in una forma canonica di base ammissibile, con il seguente tableau:

0	$4 - 5w$	$4w - 1$	0	0	0	0
$7/2$	-1	1	1	0	0	0
$11/2$	1	1	0	1	0	0
9	2	1	0	0	1	0
4	1	0	0	0	0	1

Caso $0 < w < 1/4$ Il secondo costo ridotto è negativo, e quindi occorre compiere un'operazione di pivot sull'elemento $a_{12} = 1$, facendo entrare in base x_2 e facendone uscire x_3 . Ne risulta il tableau:

$7/2(1 - 4w)$	$3 - w$	0	$1 - 4w$	0	0	0
$7/2$	-1	1	1	0	0	0
2	2	0	-1	1	0	0
$11/2$	3	0	-1	0	1	0
4	1	0	0	0	0	1

dove tutti i costi ridotti sono positivi. La soluzione ottima è $x^\circ = (0, 7/2)$ e vale $z^* = 7/2(4w - 1)$.

Caso $w = 1/4$ Questo caso limite corrisponde alla situazione in cui la forma canonica rispetto alla base x_3, x_4, x_5, x_6 è ottima, ma la variabile x_2 può entrare in base senza peggiorare la soluzione corrente, per cui anche la base x_2, x_4, x_5, x_6 è ottima. In effetti tutte le soluzioni che sono combinazione convessa di $(0, 0)$ e $(0, 7/2)$ risultano ottime. È un'infinità di punti paretiani. Si noti che il gradiente della funzione obiettivo combinata è perfettamente perpendicolare al vincolo $x_1 = 0$. Infatti, per $w = 1/4$, è $f(x) = 15/4x_1$.

Caso $1/4 < w < 4/5$ Siamo immediatamente in forma canonica di base ottima. La soluzione ottima è $x^\circ = (0, 0)$ e ha valore nullo.

Caso $w = 4/5$ Questo caso limite corrisponde alla situazione in cui la forma canonica rispetto alla base x_3, x_4, x_5, x_6 è ottima, ma la variabile x_1 può entrare in base senza peggiorare la soluzione corrente, per cui anche la base x_1, x_3, x_4, x_5 è ottima. In effetti tutte le soluzioni che sono combinazione convessa di $(0, 0)$ e $(4, 0)$ risultano ottime. È un'infinità di punti paretiani. Si noti che il gradiente della funzione obiettivo combinata è perfettamente perpendicolare al vincolo $x_2 = 0$. Infatti, per $w = 4/5$, è $f(x) = 12/5 x_2$.

Caso $4/5 < w < 1$ Il primo costo ridotto è negativo, e quindi occorre compiere un'operazione di pivot sull'elemento $a_{41} = 1$, facendo entrare in base x_1 e facendone uscire x_6 .

$4(5w - 4)$	0	$4w - 1$	0	0	0	$5w - 4$
$15/2$	0	1	1	0	0	1
$3/2$	0	1	0	1	0	-1
1	0	1	0	0	1	-2
4	1	0	0	0	0	1

Fortunatamente, ora siamo in condizioni di ottimalità (per $w > 4/5$, tutti i coefficienti di costo ridotto sono ≥ 0). La soluzione ottima è $x^\circ = (4, 0)$ e vale $z^* = 4(4 - 5w)$.

5.6.3 Analisi di sensitività

L'analisi di sensitività è in generale lo studio di come la soluzione di un problema cambia al variare di uno o più parametri numerici. Nel metodo dei pesi, essa consiste nel partizionare la regione $\{w \in \mathbb{R}^p : w_l > 0, \sum_{l=1}^p w_l = 1\}$ nei supporti delle singole soluzioni, cioè nei sottoinsiemi in corrispondenza ai quali ciascuna soluzione supportata risulta ottima per il problema a singolo obiettivo $\sum_{l=1}^p w_l f_l$.

Se il problema è finito, questi sottoinsiemi sono in numero finito. Se, in aggiunta, ha solo due obiettivi, l'insieme dei possibili vettori di peso è il segmento del piano $w_1 w_2$ compreso fra $(0, 1)$ e $(1, 0)$. In tal caso, si può rappresentare con un diagramma l'andamento dell'obiettivo per ciascuna soluzione al variare di uno dei due pesi (l'altro è fissato in modo da garantire la somma unitaria). Questo studio consente di determinare facilmente il supporto di ciascuna soluzione, che è un intervallo chiuso di valori per il peso rimanente.

Esempio 31 Riprendiamo l'Esempio 28 sul problema dell'albero ricoprente minimo. La Figura 5.11 illustra l'andamento, al variare del peso w fra 0 e 1, dell'obiettivo combinato $f_w(x) = wf_1(x) + (1 - w)f_2(x)$ per ciascuna delle tre soluzioni $x \in X$. Ricordiamo che in questo caso il problema ha due pesi, ma la condizione di normalizzazione li riconduce a uno solo. Per ogni x fissata, quindi, l'obiettivo $f_w(x)$ è una semplice funzione di una sola variabile w . Per di più, tale funzione è lineare per costruzione. Quindi per tracciarne l'andamento basta calcolarne il valore in due punti, per esempio per $w = 0$ e per $w = 1$.

Come si vede dalla figura, i tre andamenti lineari evolvono in modo tale che per valori di w bassi la soluzione paretiana identificata è C e per valori alti è A . La soluzione B , pur essendo paretiana, non viene mai identificata, in quanto ha supporto vuoto.

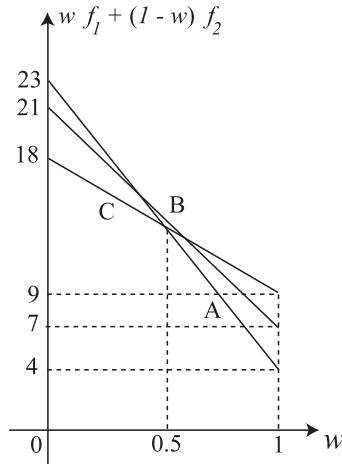


Figura 5.11: La soluzione C ha supporto $(0, 1/2]$; la soluzione A ha supporto $[1/2, 1)$; la soluzione B è globalmente paretiana, ma priva di supporto.

5.7 Metodo dei vincoli

Il metodo dei vincoli consiste nel sostituire tutti gli indicatori tranne uno con vincoli che impongano alla soluzione di rispettare una soglia di qualità e risolvere il problema ausiliario risultante. Il teorema seguente dimostra che tutte le soluzioni paretiane sono punti di ottimo per un problema di questo genere, a patto di fissare opportunamente le soglie. È un metodo molto usato e con un significato intuitivo, dato che corrisponde a conservare un solo indicatore e sostituire la ricerca di un valore ottimo per gli altri con quella di un valore soddisfacente.

Teorema 16 *Se x° è punto globalmente paretiano per gli indicatori $f_l(x)$ in X , allora per ogni indice $l^* \in \{1, \dots, p\}$, scelto a piacere, x° è punto di ottimo globale per il problema:*

$$\begin{aligned} & \min f_{l^*}(x) \\ & f_l(x) \leq \epsilon_l \quad l \in \{1, \dots, p\} \setminus \{l^*\} \\ & x \in X \end{aligned}$$

dove $\epsilon_l = f_l(x^\circ)$ per ogni $l \in \{1, \dots, p\} \setminus \{l^*\}$.

Dimostrazione. Per assurdo, supponiamo che x° non sia punto di ottimo globale per tale problema. Allora esiste un punto $x' \in X$ tale che $f_l(x') \leq \epsilon_l$ per ogni $l \neq l^*$ e $f_{l^*}(x') < f_{l^*}(x^\circ)$. Poiché $\epsilon_l = f_l(x^\circ)$, per definizione di dominanza paretiana tale punto domina x° . Siccome questo contraddice l'ipotesi, x° dev'essere punto di ottimo globale. ■

Come nel metodo dei pesi, si fa riferimento a un punto di ottimo globale per un problema ausiliario. L'idea è risolvere il problema parametricamente per tutti i valori dei coefficienti ϵ_l . Questa volta, però, la condizione sta nella tesi del teorema, e quindi è necessaria, e non sufficiente. Quindi ogni soluzione fornisce un punto candidato ad essere paretiano anziché un punto certamente paretiano; se si ha l'accortezza di calcolare ogni volta tutti i punti di ottimo globale, l'insieme dei candidati fornisce un sovrainsieme della regione paretiana. Un punto x° candidato potrebbe non essere paretiano per il problema originale. Ciò avviene quando esiste un altro punto di ottimo globale $x^{o'}$, in cui f_{l^*} ha lo stesso valore e almeno uno

degli altri indicatori ha valori migliori. Si noti che ciascun indicatore l^* fornisce in genere un sovrainsieme diverso, ed è poi possibile intersecarli per avere una stima più precisa della regione paretiana⁸. Il metodo genera comunque solo punti debolmente paretiani.

I coefficienti ϵ_l si dicono *standard* (e il metodo si chiama anche *metodo degli standard*) perché la tecnica corrisponde a concentrarsi su un obiettivo, trascurando gli altri, ma fissando per loro delle soglie di prestazione. Poiché vi sono $p - 1$ standard, tutti indipendenti tra loro, l'insieme ottenuto variandoli parametricamente è un'ipersuperficie parametrica a $p - 1$ dimensioni, esattamente come nel metodo dei pesi⁹.

Esempio 32 Si vuole posizionare una discarica di rifiuti a una distanza massima pari a 2 dalla centrale che li produce, posta nel punto $(1, 0)$. Inoltre, la discarica deve stare a nord della centrale e fuori da un parco il cui confine corre lungo la retta $x_1 + x_2 = 3$ e che giace sul versante nord-est di tale retta. Nel rispetto di tali vincoli, la discarica deve stare il più lontano possibile da una città posta all'origine degli assi e al tempo stesso dal parco.

$$\begin{aligned} \max f_1(x) &= x_1^2 + x_2^2 \\ \min f_2(x) &= x_1 + x_2 \\ g_1(x) &= (x_1 - 1)^2 + x_2^2 \leq 4 \\ g_2(x) &= x_1 + x_2 \leq 3 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

La Figura 5.12 rappresenta la regione ammissibile del problema. Dei due obiettivi, il primo mira a tenere la discarica lontana dall'origine, per cui la sua direzione di miglioramento varia da un punto all'altro ed è diretta radialmente. Il secondo, invece, migliora procedendo verso sinistra in basso.

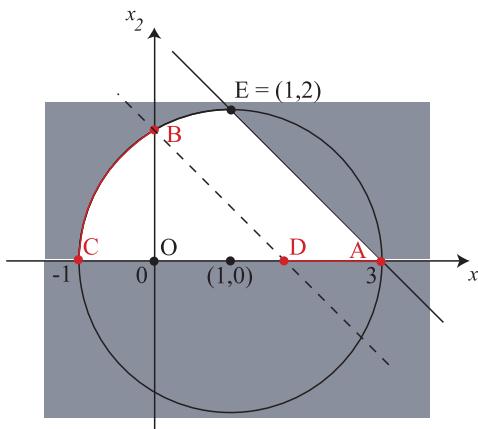


Figura 5.12: La regione ammissibile e la regione paretiana del problema

Possiamo impostare il metodo dei vincoli trasformando in vincolo l'obiettivo f_1 oppure l'obiettivo f_2 . In genere, è preferibile operare sugli obiettivi lineari, perché

⁸Sospetto che l'intersezione fornisce esattamente la regione paretiana, ma non ho sotto mano una dimostrazione, per cui mi limito a un'affermazione certamente corretta.

⁹In generale, è una superficie più ampia, ma ha lo stesso numero di dimensioni.

ne risultano vincoli più trattabili, specialmente nella risoluzione grafica. Procediamo quindi rispetto a f_2 ¹⁰.

$$\begin{aligned}\max f_1(x) &= x_1^2 + x_2^2 \\ g_1(x) &= (x_1 - 1)^2 + x_2^2 \leq 4 \\ g_2(x) &= x_1 + x_2 \leq 3 \\ g_3(x) &= -x_2 \leq 0 \\ g_4(x) &= x_1 + x_2 \leq \epsilon\end{aligned}$$

Qualsiasi metodo risolutivo è accettabile. Per esempio, si possono impostare le condizioni di KKT, ma la loro risoluzione è piuttosto ardua. Procediamo quindi graficamente. È chiaro che il vincolo parametrico $x_1 + x_2 \leq \epsilon$ corrisponde a un semispazio che va stringendo sempre più la regione ammissibile al calare di ϵ . Inoltre, la retta separatrice è parallela al confine del parco. Precisamente:

- per $\epsilon \geq 3$, il vincolo di standard non elimina alcuna soluzione ammissibile. L'ottimo globale del problema è il punto più lontano dall'origine: non è difficile vedere che tale punto è $A = (3, 0)$.
- per $\sqrt{3} < \epsilon < 3$ il vincolo di standard diventa più stringente di quello del parco ($g_2(x) \leq 0$): l'ottimo globale giace sull'asse x_1 , ed è il punto $(\epsilon, 0)$, che si sposta gradualmente da A a $D = (\sqrt{3}, 0)$.
- per $\epsilon = \sqrt{3}$ le intersezioni del vincolo di standard rispettivamente con la circonferenza $g_1 = 0$ (punto $B = (0, \sqrt{3})$) e con l'asse x_1 (punto $D = (\sqrt{3}, 0)$) sono equidistanti dall'origine, e sono entrambi ottimi globali.
- per $-1 \leq \epsilon \leq \sqrt{3}$ l'intersezione del vincolo di standard con la circonferenza $g_1(x) = 0$ diventa più lontana dall'origine rispetto all'intersezione con l'asse x_1 ; l'ottimo globale giace quindi sulla circonferenza e si sposta gradualmente da $B = (0, \sqrt{3})$ a $C = (-1, 0)$.
- per $\epsilon < -1$ non vi sono più soluzioni ammissibili.

Di conseguenza, il metodo dei vincoli indica che la regione paretiana è contenuta nell'unione del segmento AD e dell'arco di circonferenza BC . In effetti, coincide con tale insieme.

Si osservi che i punti del segmento OD sono localmente paretiani, dato che le direzioni miglioranti per entrambi gli obiettivi sono quelle volte verso il basso a destra (più in basso che a destra), e sono tutte non ammissibili.

Altri punti ancora soddisfano le condizioni di KKT, pur non essendo neppure localmente paretiani. I punti della bisettrice del primo quadrante non sono localmente paretiani (muovendosi perpendicolarmente alla bisettrice ci si allontana dall'origine mantenendo costante la distanza dal parco), ma sono ugualmente candidati, dato che i due obiettivi hanno in essi gradienti opposti e quindi $w_1 \nabla f_1 + w_2 \nabla f_2 = 0$ (i gradienti dei vincoli non compaiono perché nessun vincolo è attivo).

Esempio 33 Consideriamo lo stesso problema, ma invertendo il segno della funzione obiettivo f_2 (ad esempio, la retta $g_2(x) = 0$ rappresenta un confine di stato

¹⁰In questo caso, procedere rispetto a f_1 sarebbe probabilmente fattibile: si tratterebbe di aggiungere il vincolo $x_1^2 + x_2^2 \geq \epsilon_1$, cioè di imporre alla soluzione del problema di trovarsi all'esterno di un cerchio centrato nell'origine e avente raggio opportuno.

che non si può valicare, ma a cui ci si vuole avvicinare il più possibile, per scaricare all'estero gli effetti della discarica.

$$\begin{aligned} \max f_1(x) &= x_1^2 + x_2^2 \\ \max f_2(x) &= x_1 + x_2 \\ g_1(x) &= (x_1 - 1)^2 + x_2^2 \leq 4 \\ g_2(x) &= -x_1 - x_2 \leq -3 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

È abbastanza facile rendersi conto che questo problema ha una regione paretiana degenere, che consiste nel solo punto $A = (3, 0)$, che è contemporaneamente il più lontano dalla città e il più vicino al confine.

Risolviamo il problema trasformando in vincolo l'obiettivo f_1 (cambiandogli il segno, dato che andava massimizzato).

$$\begin{aligned} \min f_2(x) &= -x_1 - x_2 \\ g_1(x) &= (x_1 - 1)^2 + x_2^2 \leq 4 \\ g_2(x) &= -x_1 - x_2 \leq -3 \\ g_3(x) &= -x_2 \leq 0 \\ g_5(x) &= -x_1^2 - x_2^2 \leq \epsilon \end{aligned}$$

Questa volta il vincolo di standard corrisponde a richiedere una distanza minima dall'origine, ovvero a escludere i punti di un cerchio centrato nell'origine stessa:

- per $\epsilon \geq -9/2$, tutti i punti del segmento AE , con $A = (3, 0)$ ed $E = (1, 2)$ sono ammissibili per lo standard e ottimi rispetto all'obiettivo residuo. Quindi sono tutti candidati.
- per $-5 \leq \epsilon < -9/2$ la circonferenza $g_5(x) = 0$ elimina alcuni di tali punti: sono ottimi globali solo i punti dei segmenti compresi fra A e $((3 + \sqrt{-2\epsilon - 9})/2, (3 - \sqrt{-2\epsilon - 9})/2)$ e fra E e $((3 - \sqrt{-2\epsilon - 9})/2, (3 + \sqrt{-2\epsilon - 9})/2)$.
- per $-9 \leq \epsilon \leq -5$, sono ottimi globali solo i punti del segmento compreso fra A e $((3 + \sqrt{-2\epsilon - 9})/2, (3 - \sqrt{-2\epsilon - 9})/2)$.
- per $\epsilon < -9$, non vi sono più soluzioni ammissibili.

Quindi il metodo suggerisce l'intero segmento AE come candidato, benché solo il punto A sia paretiano. Vero è che, usando come obiettivo f_1 e imponendo standard su f_2 si ottiene invece la regione corretta.

5.7.1 Vantaggi e svantaggi

Il metodo dei vincoli ha diversi vantaggi:

1. è applicabile a qualsiasi problema, anche quelli dove le condizioni di KKT sono inutili (ad es., i problemi discreti);
2. è molto intuitivo, dato che spesso i decisori trovano naturale concentrarsi su un indicatore ed esprimere le proprie preferenze attraverso soglie di soddisfazione minima (si riveda nella Figura 1.2 il procedimento decisionale iterativo secondo Simon, che termina quando i decisori sono soddisfatti del risultato);

3. fornisce una sovrastima spesso abbastanza stretta della regione paretiana, che è facilmente raffinabile riapplicando il metodo con una scelta diversa dell'indicatore principale.

Al tempo stesso, però, presenta diversi svantaggi:

1. richiede di considerare tutti i valori delle soglie ϵ , le quali costituiscono un'infinità continua; questo comporta una risoluzione parametrica, che in genere non è banale, e per problemi di grandi dimensioni può diventare intrattabile;
2. per ciascun ϵ richiede di trovare tutte le soluzioni globalmente ottime (non ne basta una sola come in Programmazione Matematica);
3. nel caso dei problemi combinatorici, mentre cambiare la funzione obiettivo di solito non altera la complessità computazionale del problema (quelli polinomiali rimangono polinomiali), introdurre vincoli aggiuntivi quasi sempre aumenta la complessità (per esempio, un problema polinomiale diventa $N\mathcal{P}$ -completo).

5.7.2 Il metodo dei vincoli nella Programmazione Lineare*

Come per il metodo dei pesi, la Programmazione Lineare è un caso speciale più semplice, nel quale metodi dedicati determinano esattamente la regione paretiana. Questo richiede degli adattamenti, perché il metodo base determina ancora una sovrastima (al contrario del metodo dei pesi). Tali adattamenti, tuttavia, sono piuttosto elementari. Non trattiamo i metodi dedicati, ma ci limitiamo a mostrare un esempio di metodo parametrico esplicito, in cui si analizzano casi e sottocasi.

Esempio 34 Risolviamo nuovamente, questa volta con il metodo dei vincoli, il problema lineare considerato nell'Esempio 30.

$$\begin{aligned}
 \min f_1 &= -x_1 + 3x_2 \\
 \min f_2 &= 4x_1 - x_2 \\
 -x_1 + x_2 &\leq 7/2 \\
 x_1 + x_2 &\leq 11/2 \\
 2x_1 + x_2 &\leq 9 \\
 x_1 &\leq 4 \\
 x_1, x_2 &\geq 0
 \end{aligned}$$

La Figura 5.13 riporta la regione ammissibile del problema e le direzioni di miglioramento delle funzioni obiettivo.

Trattiamo l'obiettivo f_2 come uno standard:

$$\begin{aligned}
 \min f_1 &= -x_1 + 3x_2 \\
 -x_1 + x_2 &\leq 7/2 \\
 x_1 + x_2 &\leq 11/2 \\
 2x_1 + x_2 &\leq 9 \\
 x_1 &\leq 4 \\
 4x_1 - x_2 &\leq \epsilon \\
 x_1, x_2 &\geq 0
 \end{aligned}$$

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

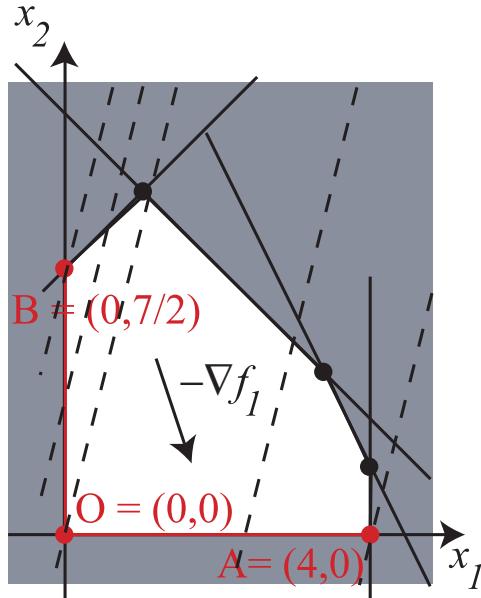


Figura 5.13: Regione ammissibile e direzioni di massimo miglioramento in un problema lineare

Siamo fortunatamente già in forma canonica di base ammissibile. Il corrispondente tableau è

0	-1	3	0	0	0	0	0
7/2	-1	1	1	0	0	0	0
11/2	1	1	0	1	0	0	0
9	2	1	0	0	1	0	0
4	1	0	0	0	0	1	0
ϵ	4	-1	0	0	0	0	1

Anzi tutto, portiamo i costi ridotti ad essere tutti non negativi: bisogna fare pivot su $a_{41} = 4$ o $a_{51} = 1$ secondo che sia $4 < \epsilon/4$ oppure $4 \geq \epsilon/4$.

Caso $\epsilon > 16$ Bisogna fare un passo di simplex su $a_{41} = 4$, ottenendo il seguente tableau.

4	0	3	0	0	0	1	0
15/2	0	1	1	0	0	1	0
3/2	0	1	0	1	0	-1	0
1	0	1	0	0	1	-1	0
4	1	0	0	0	0	1	0
$\epsilon-16$	0	-1	0	0	0	-1	1

che corrisponde a un ottimo in $x = (4, 0)$.

Caso $\epsilon \leq 16$ Si fa pivot su a_{51} , ottenendo il seguente tableau.

$\epsilon/4$	0	$11/4$	0	0	0	0	$1/4$
$7/2+\epsilon/4$	0	$3/4$	1	0	0	0	$1/4$
$11/2-\epsilon/4$	0	$5/4$	0	1	0	0	$-1/4$
$9-\epsilon/2$	0	$3/2$	0	0	1	0	$-1/2$
$4-\epsilon/4$	0	$1/4$	0	0	0	1	$-1/4$
$\epsilon/4$	1	$-1/4$	0	0	0	0	$1/4$

Se i termini noti sono tutti non negativi, siamo all'ottimo. Altrimenti, bisogna eseguire il simplesso duale. I termini noti sono tutti non negativi per $0 \leq \epsilon \leq 16$, per cui vi sono due casi.

Sottocaso $0 \leq \epsilon \leq 16$ Siamo all'ottimo, e la soluzione è $x^\circ(\epsilon) = (\epsilon/4, 0)$, vale a dire che la regione paretiana scorre lungo l'asse x_1 da A a O .

Sottocaso $\epsilon < 0$ Bisogna fare un passo di simplesso duale su $a_{52} = -1/4$, ottenendo:

3ϵ	11	0	0	0	0	0	3
$7/2+\epsilon$	3	0	1	0	0	0	1
$11/2+\epsilon$	5	0	0	1	0	0	1
$9+\epsilon$	6	0	0	0	1	0	1
4	1	0	0	0	0	1	0
$-\epsilon$	-4	1	0	0	0	0	-1

Ancora due casi: se $-7/2 \leq \epsilon < 0$, siamo all'ottimo in $x = (0, -\epsilon)$, vale a dire che la regione paretiana scorre lungo l'asse x_2 da O a B .

Se invece $\epsilon < -7/2$, dovremmo fare un ulteriore passo, ma non vi sono elementi pivot ammissibili (negativi). Questo corrisponde alla situazione di problema duale illimitato, vale a dire primale inammissibile. In effetti, il primo vincolo ha coefficienti tutti non negativi e termine noto strettamente negativo, per cui non c'è modo di soddisfarlo. Questo corrisponde al caso in cui nessuna soluzione rispetta il vincolo di standard.

In conclusione, la regione paretiana consiste nei due segmenti AO e OB , come del resto già determinato con il metodo dei pesi.

5.8 Esercizi* † ‡

Esercizio 1

Si consideri il seguente problema di programmazione matematica con due obiettivi:

$$\begin{aligned} \min f_1 &= \frac{1}{4}(x_1 - 4)^2 + \frac{1}{4}x_2^2 \\ \min f_2 &= 2 - x_2 \\ 2x_1 + x_2 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

Si tracci la regione ammissibile nel piano (x_1, x_2) e l'insieme degli impatti nel piano (f_1, f_2) .

Si determini la regione delle soluzioni paretiane per mezzo del metodo dei vincoli, spiegando via via il procedimento.

Soluzione

La regione paretiana X^* è il segmento della retta $2x_1 + x_2 = 4$ compreso fra i punti $A = (2, 0)$ e $B = (0, 4)$. La sua immagine F^* è l'arco della parabola ... compreso fra i punti $A' = (1, 2)$ e $B' = (8, -2)$.

Esercizio 2

Si consideri il seguente problema di programmazione matematica con due obiettivi:

$$\begin{aligned} \max f_1 &= x_1 - x_2 \\ \max f_2 &= x_2 \\ x_1 + x_2 &\leq 3 \\ 0 \leq x_1 &\leq 2 \\ x_2 &\geq 0 \end{aligned}$$

Si tracci la regione ammissibile nel piano (x_1, x_2) e l'insieme degli impatti nel piano (f_1, f_2) .

Si determini la regione delle soluzioni paretiane per mezzo del metodo dei vincoli, spiegando via via il procedimento.

Soluzione

La regione paretiana X^* è la spezzata poligonale ABC , con $A = (2, 0)$, $B = (2, 1)$ e $C = (0, 3)$. La sua immagine F^* è la spezzata poligonale $A'B'C'$ con $A' = (2, 0)$, $B' = (1, 1)$ e $C' = (-3, 3)$.

*Le soluzioni degli esercizi non sono state ancora riviste: eventuali segnalazioni di errore sono benvenute.

†Devo molti di questi esercizi a temi d'esame del prof. Alberto Colomni.

‡Gli esercizi che riguardano problemi finiti si possono risolvere con tutti i metodi descritti nel capitolo, ma le condizioni di KKT sono praticamente inutili. Gli esercizi che riguardano problemi infiniti a due dimensioni si possono affrontare con tutti i metodi descritti nel capitolo, salvo la definizione. Come indicato nel capitolo, i vari metodi possono fornire approssimazioni diverse della regione paretiana.

Esercizio 3

Si consideri il seguente problema di programmazione matematica con due obiettivi:

$$\begin{aligned}\min f_1 &= x'Ax + b'x \\ \min f_2 &= c'x \\ x \in X &= \{x \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}\end{aligned}$$

con

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \quad b = \begin{bmatrix} -4 \\ -8 \end{bmatrix} \quad c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Si tracci la regione ammissibile nel piano (x_1, x_2) e l'insieme degli impatti nel piano (f_1, f_2) .

Si determini la regione delle soluzioni paretiane per mezzo del metodo dei vincoli, spiegando via via il procedimento.

Soluzione

La regione paretiana X^* è la spezzata poligonale ABC , con $A = (0, 0)$, $B = (0, 1/2)$ e $C = (2, 1)$.

Esercizio 4

Si consideri il problema:

$$\begin{aligned}\max f_1 &= x_1 + 3x_2 \\ \max f_2 &= -3x_1 - 2x_2 \\ 2x_1 + x_2 &\leq 32 \\ x_1 + x_2 &\leq 20 \\ x_1 + 5x_2 &\leq 72 \\ x_1, x_2 &\geq 0\end{aligned}$$

Si determini (con il metodo dei pesi o con il metodo dei vincoli) la regione paretiana.

La si disegni nello spazio delle variabili e in quello degli obiettivi.

Soluzione

La regione paretiana X^* è la spezzata poligonale ABC , con $A = (0, 0)$, $B = (0, 72/5)$ e $C = (7, 13)$. La sua immagine F^* è la spezzata poligonale $A'B'C'$ con $A' = (0, 0)$, $B' = (216/5, -144/5)$ e $C' = (46, -47)$.

Esercizio 5

Si determini la regione Pareto-ottima, usando il metodo dei vincoli, per il seguente problema:

$$\begin{aligned}\max f_1 &= -x_1 - x_2 \\ \max f_2 &= x_1 \\ 3x_1^2 + 4x_2 &\leq 12 \\ x_2 &\geq 0\end{aligned}$$

Si disegni la regione paretiana nello spazio delle variabili e nello spazio degli obiettivi.

Soluzione*

La regione paretiana X^* è il segmento AB , con $A = (-2, 0)$ e $B = (0, 2)$. La sua immagine F^* è il segmento $A'B'$ con $A' = (?, ?)$ e $B' = (?, ?)$.

Esercise 6

Si deve localizzare un servizio di pronto intervento antincendio il più vicino possibile a un punto pericoloso nel quale è centrato un sistema di assi cartesiani. A tre chilometri a nord di tale punto scorre, in direzione est-ovest, un fiume approssimativamente rettilineo, che potrebbe essere usato per alimentare di acqua il servizio, costruendo un acquedotto dedicato, che dovrebbe essere il più corto possibile per limitare il costo. Per motivi di velocità nell'intervento, il servizio non può essere localizzato a nord del fiume, né a sudovest di una strada di grande scorrimento rappresentata dalla retta $2x_1 + x_2 = 4$.

Si determini la regione di Pareto nello spazio delle variabili e la sua immagine nello spazio degli indicatori.

Solution

In sintesi, il problema impone due vincoli alla localizzazione del servizio e due criteri di scelta: la vicinanza al punto pericoloso e la vicinanza al fiume. La regione paretiana X^* è il segmento AB , con $A = (1/2, 3)$ e $B = (8/5, 4/5)$.

Esercizio 7

Si risolva il seguente problema di programmazione a due obiettivi:

$$\begin{aligned} \max f_1 &= 9x_1^2 + 4x_2^2 - 18x_1 - 16x_2 \\ \max f_2 &= -x_1 \\ 3x_1 + x_2 &\leq 6 \\ 3x_1 + 2x_2 &\leq 9 \\ x_2 &\geq 0 \end{aligned}$$

Si disegni la regione paretiana nello spazio delle variabili e nello spazio degli obiettivi.

Soluzione

La regione paretiana X^* è la spezzata poligonale ABC , con $A = (1, 2)$, $B = (4/3, 2)$ e $C = (2, 0)$. La sua immagine F^* è una spezzata curva $A'B'C'$ con $A' = (-25, -1)$, $B' = (-24, -4/3)$ e $C' = (0, -2)$.

*Questo problema è equivalente a quello risolto nell'Esempio 20 con il metodo della trasformazione inversa.

Esercizio 8

La tabella seguente rappresenta le prestazioni di cinque alternative rispetto a quattro indicatori (tutti da massimizzare).

	a_1	a_2	a_3	a_4	a_5
f_1	100	70	60	40	20
f_2	60	45	40	100	80
f_3	60	25	20	80	100
f_4	20	100	90	50	40

Si indichino le alternative dominate (se ve ne sono) e da quali altre alternative sono dominate. Si scriva la definizione formale di dominanza paretiana.

Soluzione

L'alternativa a_3 è dominata dall'alternativa a_2 .

Esercizio 9

Si consideri il seguente problema a due obiettivi:

$$\begin{aligned} \max f_1(x) &= -x_1 + 2x_2 \\ \max f_2(x) &= 2x_1 - x_2 \\ x_1 &\leq 4 \\ x_2 &\leq 4 \\ x_1 + x_2 &\leq 7 \\ -x_1 + x_2 &\leq 3 \\ x_1 - x_2 &\leq 3 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned}$$

Si disegni la regione paretiana nello spazio delle variabili e la sua immagine in quello degli indicatori.

Soluzione

La regione paretiana X^* è la spezzata poligonale $FEDC$, con $C = (1, 4)$, $D = (3, 4)$, $E = (4, 3)$ ed $F = (4, 1)$.

Esercizio 10

Si consideri il seguente problema a due obiettivi:

$$\begin{aligned} \max f_1(x) &= x_1 - 3x_2 \\ \max f_2(x) &= -4x_1 + x_2 \\ -2x_1 + 2x_2 &\leq 7 \\ 2x_1 + 2x_2 &\leq 11 \\ x_1 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

Si disegni la regione paretiana nello spazio delle variabili e la sua immagine in quello degli indicatori.

Soluzione

La regione paretiana X^* è la spezzata poligonale AOB , con $A = (0, 7/2)$, $O = (0, 0)$ e $B = (4, 0)$.

Esercizio 11

Si consideri il seguente problema a due obiettivi:

$$\begin{aligned} \min f_1(x) &= x_1^2 + x_2^2 \\ \max f_2(x) &= x_2 \\ x_2 &\leq 10 \end{aligned}$$

Si disegni la regione paretiana nello spazio delle variabili e la sua immagine in quello degli indicatori.

Soluzione

La regione paretiana X^* è il segmento OA , con $O = (0, 0)$ e $A = (0, 10)$.

Esercizio 12

Un centro sociale cerca una nuova sede: esistono quattro alternative (A , B , C e D) oltre all'alternativa 0 (restare nella sede attuale). Si è stabilito che la scelta tra le cinque alternative debba essere definitiva e che sarà fatta in base a tre fattori: costi, accessibilità e prestigio. La tabella seguente indica i benefici associati a ciascuna alternativa e fattore in una scala tra 0 e 100.

Indicatori	A	B	C	D	0
Costi	90	90	90	1	100
Accessibilità	12	13	10	100	37
Prestigio	30	1	5	100	10

Si dica se esistono alternative dominate e da quali altre alternative sono dominate.

Soluzione

L'alternativa C è dominata dall'alternativa A , mentre l'alternativa B è dominata dall'alternativa 0.

Esercizio 13

Si consideri il seguente problema a due obiettivi:

$$\begin{aligned} \min f_1(x) &= x^2 - 4x \\ \min f_2(x) &= -x^2 \\ 0 \leq x &\leq 3 \end{aligned}$$

Si disegni la regione paretiana nello spazio delle variabili e la sua immagine in quello degli indicatori.

Soluzione

La regione paretiana X^* è l'intervallo $x \in [2; 3]$. La sua immagine F^* è l'arco della parabola ... compreso fra i punti $A' = (-4, -4)$ e $B' = (-3, -9)$.

Esercizio 14

Si consideri il seguente problema a due obiettivi:

$$\begin{aligned} \min f_1(x) &= x_1^2 + 4x_2^2 - 2x_1 - 16x_2 \\ \min f_2(x) &= -5x_1 - x_2 \\ x_1^2 + x_2^2 &\leq 8 \\ x_2 &\geq 0 \end{aligned}$$

Si descrivano analiticamente il cono ammissibile \mathcal{D}_a e il cono migliorante \mathcal{D}_m nel punto $P = (2, 2)$. In base ai due coni, può il punto P essere paretiano? Perché?

Soluzione

Il cono ammissibile e quello migliorante sono:

$$\mathcal{D}_a = \{d \in \mathbb{R}^2 : d_1 + d_2 \leq 0\} \quad \mathcal{D}_m = \{d \in \mathbb{R}^2 : d_1 \leq 0, 5d_1 + d_2 \geq 0\}$$

I due coni non hanno direzioni comuni. Quindi, P soddisfa le condizioni necessarie di KKT per la paretianità locale: può essere paretiano, anche se non è garantito che lo sia.

Esercizio 16

È dato il seguente problema decisionale con tre alternative e quattro attributi (utilità):

Attributi	<i>A</i>	<i>B</i>	<i>C</i>
u_1	0	100	80
u_2	100	83	0
u_3	70	20	100
u_4	40	100	20

Si dica se vi sono alternative dominate e da quali altre alternative sono dominate.

Soluzione

Non vi sono alternative dominate.

Esercizio 17

Si consideri il seguente problema a due obiettivi:

$$\begin{aligned} \min f_1(x) &= x_1^2 + x_2^2 - 2x_1 \\ \min f_2(x) &= -x_2 \\ x_1^2 + 4x_2^2 &\leq 8 \\ x_1 - 2x_2 &\geq 0 \end{aligned}$$

Si determini la regione paretiana nello spazio delle variabili e in quello degli obiettivi.

Soluzione

La regione paretiana X^* è la spezzata poligonale ABC , con $A = (1, 0)$, $B = (1, 1/2)$ e $C = (2, 1)$. La sua immagine F^* è la spezzata curva compresa $A'B'C'$, con $A' = (-1, 0)$, $B' = (-3/4, -1/2)$ e $C' = (1, -1)$.

Capitolo 6

Strutturazione delle preferenze

Una volta determinata la regione paretiana, una relazione di preferenza paretiana non dà altre informazioni utili per stringere ulteriormente la scelta di una soluzione. In linea di principio, il procedimento dovrebbe arrestarsi. Nella pratica, però, una decisione va presa, e quindi si deve trovare il modo di proseguire. Inoltre, la realtà è in genere più complessa dell'ipotesi paretiana secondo la quale il decisore non è affatto in grado di confrontare gli indicatori fra loro. In pratica, il decisore potrebbe essere in grado di dire qualcosa in più, oltre al fatto che, presi singolarmente, gli indicatori sono dei costi (o dei benefici). Si potrebbe quindi arricchire la relazione di preferenza, introducendo altre coppie (f, f') con $f \preceq f'$. L'ideale sarebbe arrivare a ottenere una relazione completa, e dunque un ordine totale.

A questo punto, si potrebbe anche pensare che lo studio del caso paretiano sia inutile, ma questo non è vero, perché spesso il semplice fatto di presentare al decisore la regione paretiana (come elenco di soluzioni o con una rappresentazione grafica), anziché l'intera regione ammissibile, lo aiuta a chiarirsi le idee e a raffinare la sua relazione di preferenza in un secondo tempo, arricchendola con l'aggiunta di nuove coppie confrontabili e conservando quelle definite dall'ipotesi paretiana, senza tradire ciò che davvero sente.

Nel seguito elenchiamo alcune tecniche per aggiungere la completezza a una relazione paretiana. Le prime sono elementari e piuttosto brutali, l'ultima ha invece un notevole grado di raffinatezza ed è la base della teoria economica classica, ma richiede anche ipotesi piuttosto pesanti e un procedimento molto faticoso.

6.1 Ordine lessicografico

Si chiede al decisore di ordinare totalmente gli indicatori. L'insieme non ordinato $P = \{1, \dots, p\}$ diventa quindi una *sequenza*, che per semplicità indicheremo con lo stesso simbolo:

$$P = (\pi_1, \dots, \pi_p)$$

e che rappresenterà una *gerarchia* fra indicatori. Ciò fatto, si procede come segue:

- si determinano tutte le soluzioni in cui il primo indicatore assume valore ottimo:

$$X_{\pi_1}^{\circ} = \arg \min_{x \in X} f_{\pi_1}(x)$$

Se rimane una sola soluzione ($|X_{\pi_1}^\circ| = 1$), viene scelta e si termina.

- Se rimangono più soluzioni ($|X_{\pi_1}^\circ| > 1$), si determinano fra loro quelle ottime rispetto al secondo indicatore:

$$X_{\pi_1 \pi_2}^\circ = \arg \min_{x \in X_{\pi_1}^\circ} f_{\pi_2}(x)$$

Se rimane una sola soluzione ($|X_{\pi_1 \pi_2}^\circ| = 1$), viene scelta e si termina.

- ...

- Se anche al p -esimo passo rimangono più soluzioni ($|X_{\pi_1 \dots \pi_{p-1}}^\circ| > 1$), si sceglie fra loro una qualunque soluzione ottima rispetto all'ultimo indicatore:

$$x_{\pi_1 \dots \pi_p}^\circ \in \arg \min_{x \in X_{\pi_1 \dots \pi_{p-1}}^\circ} f_{\pi_p}(x)$$

Osservazione 4 La soluzione generata dal metodo dell'ordine lessicografico è per costruzione certamente paretiana, qualunque sia l'ordine scelto.

Dimostrazione. La proprietà si può dimostrare per assurdo: un'ipotetica soluzione dominata comporterebbe l'esistenza di una soluzione migliore rispetto ad almeno un indicatore ed equivalente rispetto agli altri. Tale soluzione passerebbe indenne i filtri rispetto a questi indicatori, ma prima o poi risulterebbe migliore di quella generata dal metodo, contraddicendo il procedimento. ■

Si noti che, però, il metodo richiede nei primi $p - 1$ passi di trovare tutte le soluzioni ottime, e non una sola, come di solito avviene in Programmazione Matematica. Violare tale requisito comporterebbe la possibilità di ottenere soluzioni dominate. Questo rende il metodo potenzialmente impegnativo dal punto di vista computazionale, ma in pratica basta considerare i primi due o tre indicatori per isolare una singola soluzione.

Se la regione paretiana X° è nota, ovviamente conviene applicare il metodo partendo da essa, anziché dall'intera regione ammissibile X . Questo accelera il metodo e garantisce che la soluzione finale sia paretiana.

L'ipotesi che il decisore sia in grado di fornire un ordine totale fra gli indicatori cioè che consideri ciascuno assolutamente dominante rispetto ai successivi è in generale una forzatura, ma consente di arrivare comunque a una decisione di qualche ragionevolezza. Ovviamente, la soluzione risultante dipende in maniera fondamentale dall'ordinamento scelto.

Esempio 35 Un navigatore avanzato offre la possibilità di scegliere non solo il percorso più veloce o quello più economico, ma di applicare un metodo lessicografico di ordinamento fra i due indicatori (tempo e costo). Se l'utente indica l'ordine (costo, tempo), il navigatore enumera tutti i percorsi di costo minimo (ad es., tre percorsi da 10 euro, uno da 24 ore, uno da 36 ore e uno da 28 ore). Quindi, fra questi percorsi sceglie uno di quelli che richiedono il tempo minimo (in questo caso, il percorso ottimo è unico e richiede 24 ore). Percorsi poco più costosi (ad es., 15 o 20 euro) vengono ignorati anche se molto più brevi (ad es., 2 ore).

6.2 Ordine lessicografico con livelli di aspirazione

L'idea che un indicatore sia assolutamente dominante rispetto agli altri è spesso troppo forzata, nel senso che porta a soluzioni molto sbilanciate: ottime rispetto al

primo indicatore, ma pessime rispetto agli indicatori successivi. Il decisore umano spesso non cerca l'ottimalità, ma un livello soddisfacente di prestazione (Herbert Simon chiamava tali soluzioni “satisficing” da “satisfy” e “sufficing”). Questa linea di pensiero porta a modificare il metodo dell'ordinamento lessicografico chiedendo al decisore di ordinare totalmente gli indicatori, ma anche di indicare un *livello di aspirazione* ϵ_{π_l} per ognuno tranne il primo ($l = 2, \dots, p$). Si determina così una regione ammissibile ristretta

$$X_\epsilon = \{x \in X : f_{\pi_l}(x) \leq \epsilon_{\pi_l} \text{ per } l = 2, \dots, p\}$$

Per il primo indicatore non è necessario fissare un livello di aspirazione, dato che esso verrà ancora semplicemente minimizzato. Qualora la regione ammissibile ristretta X_ϵ fosse vuota, si dovrà chiedere al decisore di rilassare uno o più dei suoi livelli di aspirazione.

Ciò fatto, si applica il procedimento descritto per il metodo dell'ordine lessicografico, ma operando sempre nella regione ammissibile ristretta:

- si determinano tutte le soluzioni in cui il primo indicatore assume valore ottimo e i successivi assumono valori non peggiori dei corrispondenti livelli di aspirazione:

$$X_{\epsilon, \pi_1}^\circ = \arg \min_{x \in X_\epsilon} f_{\pi_1}(x)$$

Se rimane una sola soluzione ($|X_{\epsilon, \pi_1}^\circ| = 1$), viene scelta e si termina.

- Se rimangono più soluzioni ($|X_{\epsilon, \pi_1}^\circ| > 1$), si determinano fra loro quelle ottime rispetto al secondo indicatore:

$$X_{\epsilon, \pi_1 \pi_2}^\circ = \arg \min_{x \in X_{\epsilon, \pi_1}} f_{\pi_2}(x)$$

Se rimane una sola soluzione ($|X_{\epsilon, \pi_1 \pi_2}^\circ| = 1$), viene scelta e si termina.

- ...

- Se anche al p -esimo passo rimangono più soluzioni ($|X_{\epsilon, \pi_1 \dots \pi_{p-1}}^\circ| > 1$), si sceglie fra loro una qualunque soluzione ottima rispetto al p -esimo indicatore:

$$x_{\epsilon, \pi_1 \dots \pi_p}^\circ \in \arg \min_{x \in X_{\epsilon, \pi_1 \dots \pi_{p-1}}^\circ} f_{\pi_p}(x)$$

L'idea alla base di questo metodo è che ogni indicatore sia dominante rispetto ai successivi solo finché i loro livelli di aspirazione sono tutti rispettati. Nel momento in cui uno di loro tocca il livello di aspirazione, diviene più importante ed è vietato peggiorarlo ulteriormente.

Si noti la parentela che esiste fra questo metodo e il metodo dei vincoli per determinare la regione paretiana: i livelli di aspirazione corrispondono a una particolare scelta degli standard usati nel metodo dei vincoli. La differenza principale è che gli standard ϵ_l vengono fatti variare in tutti i modi possibili, anziché essere fissati una volta per tutte al principio. Va detto, però, che il decisore potrebbe anche fare degli esperimenti, modificando la scelta dei livelli di aspirazione in base ai risultati che producono. Questo rende il metodo ancora più vicino a una versione del metodo dei vincoli basata sul campionamento. Questa relazione fra i due metodi suggerisce che la soluzione generata sia quanto meno candidata ad essere paretiana; in realtà, il fatto di ottimizzare via via gli indicatori garantisce che lo sia sempre.

Osservazione 5 La soluzione generata dal metodo dell'ordine lessicografico con livelli di aspirazione è per costruzione certamente paretiana, qualunque sia l'ordine scelto e il valore dei livelli di aspirazione, purché producano una regione ammissibile ristretta non vuota.

Dimostrazione. Anche in questo caso, la proprietà si può dimostrare per assurdo, ipotizzando la generazione di una soluzione dominata, che però rispetterebbe gli stessi livelli di aspirazione della soluzione che la domina e al tempo stesso sarebbe peggiore per uno o più indicatori, contraddicendo il procedimento. ■

Ovviamente, la soluzione risultante dipende dall'ordinamento scelto, ma anche dai $p - 1$ livelli di aspirazione ϵ_{π_l} (per $l = 2, \dots, p$). Siccome sono tutte scelte potenzialmente arbitrarie, si tratta ancora di una forzatura, che consente di arrivare a una decisione di qualche ragionevolezza in modo semplice. I livelli di aspirazione, però, consentono di rendere la soluzione finale meno sbilanciata a favore del primo indicatore.

Esempio 36 Un navigatore avanzato offre la possibilità di applicare un metodo lessicografico di ordinamento con livelli di aspirazione. L'utente indica l'ordine (tempo, costo), e una soglia pari a $\epsilon = 20$ euro sul costo (il metodo richiede $p - 1$ soglie, in questo caso una sola, perché il primo indicatore viene ottimizzato). Il navigatore esclude quindi tutti i percorsi di costo superiore a 20 euro (ad es., rimangono i tre percorsi da 10 euro visti nell'Esempio 35, cioè quelli da 24, 36 e 28 ore, e i percorsi da 15 e 20 euro visti prima). Quindi, fra i percorsi residui il navigatore enumera quelli che richiedono il tempo minimo: due percorsi richiedono 2 ore (e, rispettivamente, 15 e 20 euro). A questo punto, il navigatore minimizza il costo e sceglie il percorso da 2 ore e 15 euro. La soglia di 20 euro è ovviamente arbitraria e fa ignorare percorsi anche di poco più costosi che potrebbero essere molto più brevi.

6.3 Punto utopia

Ciascuno degli indicatori ha un valore ottimo, ottenuto ignorando completamente gli altri indicatori

$$f_l^\circ = \min_{x \in X} f_l(x)$$

La determinazione di questo valore è un problema di Programmazione Matematica, a volte computazionalmente difficile, ma in generale possibile.

Il vettore costituito dai valori ottimi di tutti i p indicatori costituisce un impatto $f^U = [f_1^\circ \dots f_p^\circ]^T$. Se $f^U \in F$, esso è necessariamente l'unica soluzione paretiana, e quindi determina la soluzione da scegliere (si veda l'Esempio). In generale, però tale impatto non sta in F , cioè non è realizzabile scegliendo una soluzione ammissibile. Esso viene detto *punto utopia*, perché costituisce un ideale irraggiungibile.

Un'idea euristica per scegliere una soluzione è di determinare una di quelle che hanno l'impatto più vicino al punto utopia. Questa definizione è apparentemente innocua e ragionevole. Tuttavia, si basa su una nozione di *distanza* nello spazio degli indicatori che non è affatto ovvia. Le distanze più frequentemente usate sono:

- la distanza di Manhattan (L_1):

$$d(f, f') = \sum_{l \in P} |f_l - f'_l|$$

- la distanza euclidea (L_2):

$$d(f, f') = \sqrt{\sum_{l \in P} (f_l - f'_l)^2}$$

- la massima delle distanze relative ai singoli attributi (L_∞):

$$d(f, f') = \max_{l \in P} |f_l - f'_l|$$

ma se ne possono definire infinite (vedi Figura 6.1).

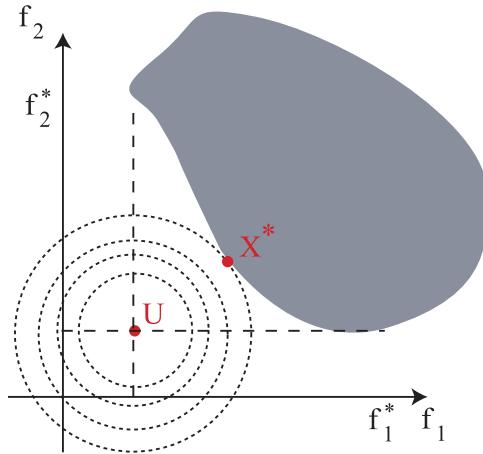


Figura 6.1: Punto utopia e soluzione ammissibile più vicina secondo la metrica L_2

Inoltre, la maggior parte delle definizioni di distanza combina in una sola grandezza valori associati a diversi indicatori, espressi in unità di misura disomogenee. Prima di poterli combinare, occorre standardizzarli moltiplicandoli per opportuni coefficienti. La scelta dei coefficienti è complessa, e almeno in parte arbitraria.

6.4 Teoria dell'utilità a molti attributi

La teoria dell'utilità a molti attributi¹ assume che la relazione di preferenza del decisore sia effettivamente un ordine debole, dotato di una funzione valore conforme, ma che il decisore non sia in grado di esplicitare la funzione valore senza aiuto. Si pone quindi il problema di ricavare dalla relazione di preferenza II la funzione valore conforme $u : F \rightarrow \mathbb{R}$. Nel seguito, adotteremo il gergo economico, chiamando $u(f)$ *funzione di utilità*.

In un primo tempo, affronteremo il problema nella sua forma più generale, presentando uno strumento grafico e una procedura che in teoria dovrebbero consentire di risolverlo, ma questa procedura si rivelerà molto complessa, costosa e soggetta a errori. Ci concentreremo quindi su una famiglia specifica di funzioni di utilità, le cui caratteristiche rendono più semplice (anche se comunque non elementare) ricostruirle a partire dalle preferenze espresse dal decisore. Discuteremo quindi delle condizioni necessarie e sufficienti per poter affermare che la relazione di preferenza sia esprimibile attraverso una di tali funzioni e infine descriveremo una procedura per ricostruire la funzione stessa.

¹In inglese, *Multi Attribute Utility Theory*, ovvero MAUT.

6.4.1 Curve di indifferenza

Definizione 21 Dato un insieme di impatti F nello spazio degli indicatori \mathbb{R}^p e una relazione di preferenza Π , si definisce curva di indifferenza ogni sottoinsieme di impatti fra loro indifferenti.

Sotto opportune ipotesi di continuità e regolarità, impatti vicini sono simili anche dal punto di vista della preferenza². Ogni sottoinsieme di impatti fra loro indifferenti costituisce un'ipersuperficie regolare in \mathbb{R}^p . Siccome le curve di indifferenza sono infinite, una per ogni valore reale $u(f)$, in generale ogni curva è un'ipersuperficie a $p - 1$ dimensioni. Le varie curve sono legate da una relazione di ordine totale; infatti, non ci sono impatti incomparabili e gli impatti indifferenti giacciono tutti sulla stessa curva, per cui impatti su curve diverse sono legati da una preferenza stretta, e tutti gli impatti di una delle due curve sono strettamente preferibili a tutti quelli dell'altra. Infine, la completezza della relazione garantisce che la famiglia delle curve di indifferenza nel complesso copre l'intero insieme F (vedi Figura 6.2).

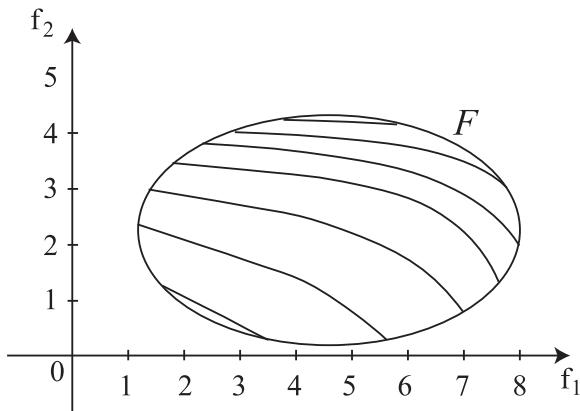


Figura 6.2: Curve di indifferenza: vi sono ∞ curve totalmente ordinate che coprono complessivamente tutto l'insieme F ; ogni curva contiene ∞^{p-1} impatti indifferenti fra loro.

Curve di indifferenza e funzione di utilità

Dal Teorema 3 della Sezione 3.5.6 sappiamo che, se una relazione di preferenza ammette una funzione valore conforme, ne ammette infinite. Le curve di indifferenza di queste funzioni sono però sempre le stesse.

Osservazione 6 Tutte le funzioni valore u conformi a una relazione di preferenza Π hanno le stesse curve di indifferenza $u(f) = c$, dove $c \in \mathbb{R}$ è un parametro reale che identifica ciascuna curva. Ogni funzione valore u associa valori diversi di c a ogni curva, ma tali valori sono ordinati allo stesso modo.

Esempio 37 Consideriamo le funzioni di utilità $u(f_1, f_2) = f_1^2 f_2^3$ e $u'(f_1, f_2) = 2 \log f_1 + 3 \log f_2$. Gli indicatori f_1 e f_2 usati in questo esempio esprimono un

²Non approfondiremo queste ipotesi: in generale, esse richiedono un insieme di impatti F continuo. Molti dei concetti che seguono, infatti, non hanno senso per problemi discreti. In questi problemi, occorre estendere gli impatti, prolungando gli indicatori che assume solo valori discreti (per esempio, "caramelle") a valori reali con una qualche forma di interpolazione (per esempio, "porzioni di caramella"). Le curve di indifferenza vengono definite su questo insieme di impatti esteso, e poi limitate all'insieme discreto originale.

beneficio, ma il concetto vale per qualsiasi tipo di indicatore. Le due funzioni sono equivalenti, perché legate dalla trasformazione invertibile $u' = \log u$. La Figura 6.3 riporta le relative curve di indifferenza: esse coincidono perfettamente, ma la stessa curva nei due casi è associata a valori diversi di utilità.

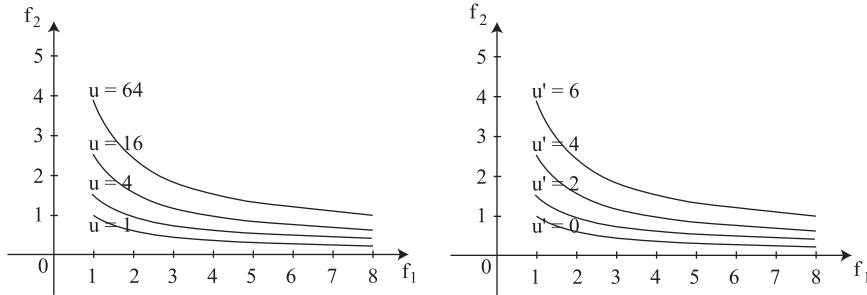


Figura 6.3: Due funzioni utilità diverse, ma equivalenti: $u(f_1, f_2) = f_1^2 f_2^3$ e $u'(f_1, f_2) = 2 \log f_1 + 3 \log f_2$. Le curve di indifferenza coincidono, pur corrispondendo a valori diversi.

6.4.2 Determinare la funzione di utilità

Le curve di indifferenza aiutano a stimare una funzione valore conforme alla relazione di preferenza attraverso il seguente procedimento:

1. si propone al decisore un campione $\tilde{F} \subseteq F$ di impatti, chiedendogli di confrontarli a coppie;
2. in base ai confronti, si classificano gli impatti:
 - gli impatti fra loro indifferenti si raccolgono in classi di equivalenza;
 - le classi di equivalenza si ordinano in base alla preferenza;
3. basandosi sulla forma delle curve di indifferenza nello spazio degli indicatori, si ipotizza una famiglia di funzioni di utilità $u_\alpha(f_1, \dots, f_p)$, definita a meno di un vettore di parametri numerici α ;
4. si determinano i valori dei parametri α_r ($r = 1, \dots, s$) imponendo agli impatti indifferenti fra loro di assumere lo stesso valore di utilità;
5. si compiono verifiche di coerenza, generando nuovi impatti, ricavando con la funzione di utilità stimata le preferenze fra tali impatti e chiedendone conferma al decisore;
6. in caso di errori, si modifica la famiglia di funzioni di utilità e si ripete il procedimento.

Il procedimento è chiaramente molto delicato. Se si considera un campione piccolo, è molto probabile che le curve stimate siano errate; se il campione è grande, il carico di lavoro richiesto al decisore diventa rapidamente enorme (se si valutano k valori diversi per ogni indicatore, il campione contiene k^p impatti diversi, un numero che diventa rapidamente intrattabile). Inoltre, l'arbitrarietà nella scelta della famiglia di funzioni di utilità apre la strada a lunghi cicli di tentativi ed errori.

L'unica possibilità è che la relazione di preferenza del decisore sia molto semplice, e quindi le curve di indifferenza abbiano una forma analitica molto semplice,

dipendente da pochi parametri, facili da stimare. Siamo quindi interessati a valutare casi particolari che consentono una facile stima, nella speranza che siano abbastanza vicini al caso concreto da poter essere usati come approssimazioni della realtà. I corsi di economia discutono parecchie famiglie di funzioni di utilità, che corrispondono a diverse ipotesi sulla relazione di preferenza del decisore. Proprietà che vengono spesso postulate sono:

1. *invertibilità*: per ogni valore fissato di $p - 1$ indicatori, esiste un solo valore dell'indicatore residuo che produce una data utilità; sotto questa ipotesi, una curva di indifferenza $u(f) = c$ si può scrivere anche in forma esplicita come $f_l = f_l(c, f_1, \dots, f_p)$;
2. *monotonia*: per compensare variazioni di un indicatore, gli altri devono variare in un senso ben determinato; per esempio, quando tutti gli indicatori sono costi, per compensare la crescita di uno occorre il calo di un altro, e quindi le curve sono decrescenti (lo stesso avviene se tutti gli indicatori sono benefici).
3. *convessità* (o *concavità*): le curve di indifferenza compensano la crescita di un indicatore di una data quantità con variazioni degli altri che crescono (o calano) col valore del primo indicatore; l'idea alla base di questa ipotesi è che gli indicatori rappresentino risorse: se una risorsa è scarsa, aumentarla porta una grande utilità, compensata da un calo forte di altre risorse; se invece è abbondante, aumentarla della stessa quantità porta un'utilità piccola, compensata da un debole calo di altre risorse; in questo caso, le curve sono convesse (nel caso di indicatori che esprimono costi, sarebbero invece concave).

Esempio 38 *Si vuole decidere in quale città andare a vivere sulla base (molto semplificata) di due indicatori: il livello di inquinamento della zona e il livello della tassazione locale. Due città con livelli di inquinamento diversi saranno indifferenti solo se la più inquinata ha un livello di tasse inferiore. Quindi le curve di indifferenza sono monotone decrescenti e invertibili. Inoltre, se il livello di inquinamento in una città è basso, un aumento di valore fissato è compensabile con una piccola riduzione delle tasse, mentre se è alto, lo stesso aumento richiederà un forte incentivo fiscale.*

Esempio 39 *Una famiglia di funzioni di utilità molto usate in economia sono le funzioni di Cobb-Douglas, definite come:*

$$u_\alpha(f_1, \dots, f_p) = \prod_{l=1}^p f_l^{\alpha_l}$$

Queste funzioni considerano gli indicatori f_l come benefici (sono infatti state proposte da economisti), richiedono un parametro per ogni indicatore e hanno la proprietà di dar luogo a curve di indifferenza che somigliano a degli iperbolidi. La funzione di utilità $u(f_1, f_2) = f_1^2 f_2^3$, le cui curve di indifferenza sono riportate nella Figura 6.3, appartiene a questa famiglia: i suoi parametri sono infatti $\alpha_1 = 2$ e $\alpha_2 = 3$. Queste funzioni sono invertibili ($f_1 = \sqrt{c/f_2^3}$ e $f_2 = \sqrt[3]{c/f_1^2}$ per $f_1 \geq 0$ e $f_2 \geq 0$), monotone decrescenti in senso stretto (si calcoli la derivata prima) e convesse (si calcoli la derivata seconda).

6.4.3 Funzioni di utilità additive

Esistono funzioni di utilità che si prestano più facilmente ad essere stimate a partire da un campionamento della relazione di preferenza sugli impatti.

Definizione 22 Definiamo una funzione di utilità additiva quando è esprimibile come somma di funzioni dei singoli indicatori:

$$u(f_1, \dots, f_p) = \sum_{l=1}^p u_l(f_l)$$

Il fatto che per ogni funzione di utilità ne esistono infinite equivalenti rende complicato valutare se esista o no una funzione additiva nella situazione corrente. Un classico esempio sono le funzioni di Cobb-Douglas che di per sé non sono additive, ma, come già osservato nell'esempio di Figura 6.3, ammettono funzioni equivalenti che sono additive:

$$u'(f) = \log u(f) = \log \prod_{l=1}^p f_l^{\alpha_l} = \sum_{l=1}^p \alpha_l \log f_l = \sum_{l=1}^p u_l(f_l)$$

dove $u_l(f_l) = \alpha_l \log f_l$ per ogni $l \in P$.

Se la funzione di utilità è additiva, il problema di stimarla viene ricondotto alla stima di p funzioni di una sola variabile. Se si richiamano alla mente i casi di studio del Capitolo 56, ci si rende conto che il numero degli indicatori può essere enorme, e che la loro natura può essere estremamente eterogenea, tanto da rendere ridicola la pretesa di confrontare impatti estratti da punti molto lontani dell'insieme F . Se si potesse trattare con funzioni di una sola variabile, il compito di stimare l'utilità sarebbe molto più facile; in particolare, si potrebbe affidare la stima di ciascuna funzione a un diverso esperto di settore. Ovviamente, esistono situazioni intermedie, in cui la funzione di utilità non è completamente additiva, ma quanto meno scomponibile in una somma di funzioni che interessano blocchi di pochi indicatori fra loro indipendenti. Sarebbe comunque un passo avanti.

Siccome a priori non conosciamo la funzione di utilità, ma la relazione di preferenza, dobbiamo indagare quali proprietà della relazione di preferenza consentano di garantire l'esistenza di una funzione di utilità additiva.

6.4.4 Indipendenza preferenziale

Dato l'insieme degli attributi $P = \{1, \dots, p\}$, l'indipendenza preferenziale è una proprietà di un sottoinsieme di attributi $L \subset P$ rispetto al proprio complemento $\bar{L} = P \setminus L$. Se per semplicità riordiniamo gli attributi raccogliendo nelle prime posizioni quelli del sottoinsieme L , possiamo scrivere che $f = \begin{bmatrix} f_L \\ f_{\bar{L}} \end{bmatrix}$, dove f_L e $f_{\bar{L}}$ sono i vettori dell'impatto f corrispondenti, rispettivamente, agli indicatori di L e di \bar{L} .

Definizione 23 Un sottoinsieme proprio di indicatori $L \subset P$ è preferenzialmente indipendente dal sottoinsieme complementare \bar{L} quando, dati due impatti qualsiasi con valori identici per gli indicatori in \bar{L} , la relazione di preferenza fra loro non dipende da tali valori:

$$\begin{bmatrix} f_L \\ \phi \end{bmatrix} \preceq \begin{bmatrix} f'_L \\ \phi \end{bmatrix} \Leftrightarrow \begin{bmatrix} f_L \\ \psi \end{bmatrix} \preceq \begin{bmatrix} f'_L \\ \psi \end{bmatrix}$$

per ogni f_L, f'_L, ϕ, ψ tale che

$$\begin{bmatrix} f_L \\ \phi \end{bmatrix}, \begin{bmatrix} f'_L \\ \phi \end{bmatrix}, \begin{bmatrix} f_L \\ \psi \end{bmatrix}, \begin{bmatrix} f'_L \\ \psi \end{bmatrix} \in F$$

Esempio 40 Si supponga di dover scegliere il menù di un pranzo veloce, abbinando una portata e un vino. Le portate disponibili sono: brasato, arrosto, polpette, salmone, pesce spada; i vini disponibili sono: Barolo, Nebbiolo, Erbaluce, Arneis. Abbiamo quindi un insieme X fatto di 20 soluzioni ammissibili (le combinazioni di 5 portate e 4 vini). Il decisore è abbastanza rozzo da percepire solo la distinzione fra carne e pesce e quella fra vino rosso e bianco. Quindi l'insieme degli impatti si compone di 4 elementi: $F = (\text{carne, rosso}), (\text{carne, bianco}), (\text{pesce, rosso}), (\text{pesce, bianco})$. Supponendo che segua l'uso tradizionale di abbinare piatti di carne con il vino rosso e piatti di pesce con il vino bianco, gli indicatori “tipo di portata” e “tipo di vino” non sono preferenzialmente indipendenti. Infatti, fissato il vino rosso, le portate di carne sono preferibili a quelle di pesce; fissato il vino bianco, le portate di pesce sono preferibili a quelle di carne. Lo stesso vale nel caso inverso: fissata la carne, è preferibile il vino rosso; fissato il pesce, lo è il vino bianco.

Esempio 41 Si supponga di dover definire un processo industriale per la produzione di pentole e coperchi, valutandolo rispetto agli attributi “costo”, “numero di pentole” e “numero di coperchi”. I processi preferibili hanno costo inferiore e numero di pentole e coperchi il più possibile vicini fra loro. La coppia (“costo”, “numero di pentole”) non è indipendente dall'attributo “numero di coperchi”, dato che per ogni fissato numero di coperchi è preferibile che il numero di pentole uguagli quello dei coperchi. Così pure, la coppia (“costo”, “numero di coperchi”) non è indipendente dall'attributo “numero di pentole”. Invece, la coppia (“numero di pentole”, “numero di coperchi”) è indipendente dall'attributo “costo”, perché ad ogni costo fissato la relazione di preferenza fra diverse combinazioni di pentole e coperchi risulta sempre la stessa.

Si potrebbe pensare che l'indipendenza sia una proprietà simmetrica, cioè che quando L è indipendente da \bar{L} , anche \bar{L} sia indipendente da L . L'esempio seguente mostra che non è così.

Esempio 42 Dato l'insieme degli impatti $F = \{f \in \mathbb{R}^2 : f_1 \geq 0, f_2 \geq 1\}$ e la funzione di utilità $u(f) = (f_1 - 5)f_2$, l'indicatore f_1 è preferenzialmente indipendente dall'indicatore f_2 perché, fissato $f_2 = \bar{f}_2 \geq 1$, è sempre

$$(f_1, \bar{f}_2) \preceq (f'_1, \bar{f}_2) \text{ per } f_1 \geq f'_1$$

Al contrario, f_2 dipende preferenzialmente da f_1 perché, fissato $f_1 = \bar{f}_1 \geq 1$:

$$\begin{cases} \text{quando } 0 \leq \bar{f}_1 < 5, & (\bar{f}_1, f_2) \preceq (\bar{f}_1, f'_2) \text{ per } f_2 \leq f'_2 \\ \text{quando } \bar{f}_1 = 5, & (\bar{f}_1, f_2) \sim (\bar{f}_1, f'_2) \text{ per ogni } f_2, f'_2 \\ \text{quando } \bar{f}_1 > 5, & (\bar{f}_1, f_2) \preceq (\bar{f}_1, f'_2) \text{ per } f_2 \geq f'_2 \end{cases}$$

Si potrebbe anche pensare che quando ciascun indicatore è indipendente dagli altri, tutti i sottoinsiemi di indicatori siano fra loro indipendenti. Anche questo non è vero, come mostra l'esempio seguente.

Esempio 43 Dato l'insieme degli impatti $F = \{f \in \mathbb{R}^3 : f_1 \geq 0, f_2 \geq 0, f_3 \geq 1\}$ e la funzione di utilità $u(f) = 1 / [(f_1 + f_3)(f_2 + f_3)]$, ciascun attributo è indipendente dagli altri: sono tutti costi. D'altra parte, la coppia di attributi (f_1, f_2) dipende preferenzialmente dall'attributo f_3 . Infatti, risulta che

$$\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \succ \begin{bmatrix} 4 \\ 1/2 \\ 1 \end{bmatrix} \text{ poiché } u(1, 3, 1) = \frac{1}{8} < \frac{2}{15} = u(4, 1/2, 1)$$

$$\begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} \prec \begin{bmatrix} 4 \\ 1/2 \\ 3 \end{bmatrix} \text{ poiché } u(1, 3, 3) = \frac{1}{24} > \frac{2}{49} = u(4, 1/2, 3)$$

6.4.5 Indipendenza preferenziale e additività

Indipendenza preferenziale della relazione Π e additività della funzione $u(f)$ sono strettamente legate, anche se non rigorosamente equivalenti.

Definizione 24 Si dice che un problema decisionale gode della mutua indipendenza preferenziale quando ogni sottoinsieme di indicatori $L \subset P$ è preferenzialmente indipendente dal proprio complemento \bar{L} .

La mutua indipendenza preferenziale è condizione necessaria per l'additività.

Teorema 17 Se una relazione di preferenza ammette una funzione utilità additiva, allora gode della mutua indipendenza preferenziale.

Dimostrazione. È sufficiente applicare la definizione. Per qualsiasi quaterna di vettori f_L, f'_L, ϕ, ψ che producano impatti validi:

$$\begin{bmatrix} f_L \\ \phi \end{bmatrix} \preceq \begin{bmatrix} f'_L \\ \phi \end{bmatrix} \Rightarrow u(f_L, \phi) \geq u(f'_L, \phi)$$

da cui per additività

$$\sum_{l \in L} u_l(f_l) + \sum_{l \in \bar{L}} u_l(\phi_l) \geq \sum_{l \in L} u_l(f'_l) + \sum_{l \in \bar{L}} u_l(\phi_l)$$

dove si possono sostituire le due sommatorie identiche nei due membri con altre due sommatorie identiche

$$\begin{aligned} \sum_{l \in L} u_l(f_l) + \sum_{l \in \bar{L}} u_l(\psi_l) &\geq \sum_{l \in L} u_l(f'_l) + \sum_{l \in \bar{L}} u_l(\psi_l) \Rightarrow \\ \Rightarrow \begin{bmatrix} f_L \\ \psi \end{bmatrix} &\preceq \begin{bmatrix} f'_L \\ \psi \end{bmatrix} \end{aligned}$$

■

A questo punto, è opportuno porsi due domande:

1. la mutua indipendenza preferenziale è anche sufficiente per l'additività?
2. per verificare la mutua indipendenza preferenziale è strettamente necessario verificare l'indipendenza di ogni sottoinsieme L ?

Partiamo dalla seconda domanda. Quando L contiene un solo indicatore, spesso è facile dimostrare l'indipendenza. Per esempio, quando l'indicatore rappresenta un costo o un beneficio, esso è preferenzialmente indipendente dagli altri, dato che la preferenza andrà sempre all'impatto con costo minore (o beneficio maggiore), se tutti gli altri indicatori rimangono invariati, qualunque sia il loro valore. Purtroppo, anche quando tutti gli indicatori presi singolarmente sono indipendenti, non è detto che ogni loro sottoinsieme L sia indipendente dal complemento. Fortunatamente, non è necessario considerare tutti i sottoinsiemi.

Teorema 18 Un problema decisionale con $p \geq 3$ indicatori gode della mutua indipendenza preferenziale se e solo se esiste un indice $\bar{l} \in P$ tale che ogni coppia $\{\bar{l}, l\}$ con $l \neq \bar{l}$ sia preferenzialmente indipendente dal proprio complemento $P \setminus \{\bar{l}, l\}$.

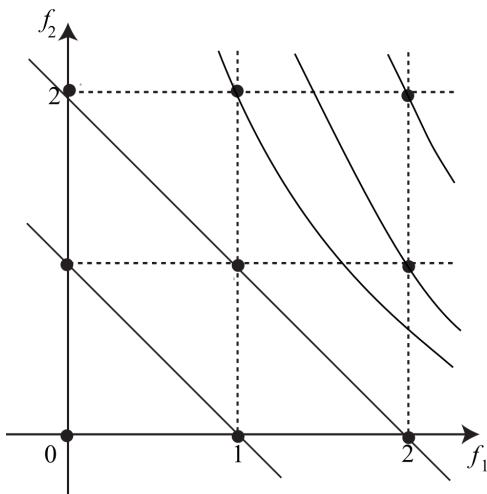


Figura 6.4: Un problema decisionale con $p = 2$ indicatori che sono preferenzialmente indipendenti, ma senza funzioni utilità additive

Quindi, per verificare la mutua indipendenza preferenziale non occorre considerare tutti i $2^p - 2$ sottoinsiemi propri di indicatori, ma solo le $p - 1$ coppie che un indicatore qualsiasi forma con gli altri: alla prima coppia non indipendente, la verifica si arresta con un risultato negativo; se tutte sono indipendenti, la verifica termina con un risultato positivo.

Teorema 19 *Un problema decisionale con $p \geq 3$ indicatori che gode della mutua indipendenza preferenziale ammette una funzione utilità $u(f)$ additiva.*

Quindi, per problemi con almeno tre indicatori, la mutua indipendenza preferenziale è condizione necessaria e sufficiente per l'additività. Purtroppo, nel caso di due soli indicatori, questo non vale.

Esempio 44 La Figura 6.4 riporta le curve di indifferenza di un problema decisionale con $p = 2$ indicatori che rappresentano dei benefici. Vale quindi la mutua indipendenza preferenziale, dato che per qualsiasi valore fissato di f_2 è preferibile che f_1 cresca, e per qualsiasi valore fissato di f_1 è preferibile che f_2 cresca. Tuttavia, non esistono funzioni di utilità additive capaci di determinare quelle curve di indifferenza.

Osserviamo infatti che:

$$(2, 0) \sim (1, 1) \sim (0, 2) \quad e \quad (1, 0) \sim (0, 1)$$

Per assurdo, supponiamo che esista una funzione di utilità additiva. Siccome $(2, 0)$ e $(0, 2)$ sono impatti indifferenti, la loro utilità è la stessa; lo stesso avviene per $(1, 0)$ e $(0, 1)$:

$$\begin{cases} u(0, 2) = u(2, 0) \Rightarrow u_1(0) + u_2(2) = u_1(2) + u_2(0) \\ u(0, 1) = u(1, 0) \Rightarrow u_1(0) + u_2(1) = u_1(1) + u_2(0) \end{cases}$$

Sottraendo membro a membro, si ottiene

$$u_2(2) - u_2(1) = u_1(2) - u_1(1)$$

e quindi

$$u_2(2) + u_1(1) = u_1(2) + u_2(1) \Rightarrow u(1, 2) = u(2, 1) \Rightarrow (1, 2) \sim (2, 1)$$

ma i due impatti $(1, 2)$ e $(2, 1)$ non sono indifferenti.

L'Esempio 44 e la Figura 6.4 non suggeriscono solo che la mutua indipendenza preferenziale sia insufficiente a garantire l'additività nel caso $p = 2$, ma anche che il difetto riguardi le curve di indifferenza: la presenza di due curve rettilinee, seguite da altre curve non rettilinee (cosa perfettamente possibile, finché le curve non si intersecano) rende impossibile la costruzione di una funzione di utilità additiva.

6.4.6 Tasso di sostituzione

La condizione mancante per garantire l'additività riguarda la pendenza delle curve di indifferenza. Abbiamo quindi bisogno di una misura di pendenza.

Definizione 25 Si dice tasso marginale di sostituzione di f_1 con f_2 in un impatto f dato il limite

$$\lambda_{12}(f) = \lim_{\delta f_1 \rightarrow 0} -\frac{\delta f_2(f, \delta f_1)}{\delta f_1}$$

dove $\delta f_2(f, \delta f_1)$ è tale che: $f + \left[\begin{array}{c} \delta f_1 \\ \delta f_2(f, \delta f_1) \end{array} \right] \sim f$.

Il tasso marginale di sostituzione, cioè, è il limite, per variazioni infinitesime di f_1 , del rapporto fra le variazioni di f_2 e di f_1 che producono impatti indifferenti a f ; il limite viene cambiato di segno perché molto spesso è negativo (per esempio, quando i due indicatori sono entrambi costi o entrambi benefici). In parole povere, la definizione ipotizza di partire da un impatto f , modificare leggermente il valore dell'indicatore f_1 e determinare la corrispondente variazione dell'indicatore f_2 che permette di rimanere sulla curva di indifferenza iniziale³.

Si possono dare tre espressioni diverse, ma equivalenti, del tasso marginale di sostituzione. La prima non è particolarmente significativa, ma è utile per ricavare le altre due. Supponiamo che le curve di indifferenza siano degli archi regolari e rappresentiamole in modo parametrico:

$$\begin{cases} f_1 = f_1(\alpha) \\ f_2 = f_2(\alpha) \end{cases}$$

Le variazioni δf_1 e δf_2 usate nella definizione di $\lambda_{12}(f)$ corrispondono a una variazione $\delta\alpha$ del parametro; in questo modo, infatti, è garantito che l'impatto rimanga sulla curva di indifferenza.

$$\lambda_{12}(f) = \lim_{\delta f_1 \rightarrow 0} -\frac{\delta f_2(f, \delta f_1)}{\delta f_1} = \lim_{\delta\alpha \rightarrow 0} -\frac{f_2(\alpha + \delta\alpha) - f_2(\alpha)}{f_1(\alpha + \delta\alpha) - f_1(\alpha)} = -\frac{\frac{df_2}{d\alpha}}{\frac{df_1}{d\alpha}} \quad (6.1)$$

Osservazione 7 Dall'Equazione (6.1) deriva immediatamente che i tassi di sostituzione marginali di f_1 rispetto a f_2 e di f_2 rispetto a f_1 sono reciproci l'uno

³Per problemi con più di due indicatori, si definisce un tasso di sostituzione $\lambda_{ll'}(f)$ per ogni coppia di indicatori (l, l') , che viene calcolato ipotizzando che tutti gli altri indicatori rimangano costanti mentre si fanno variare f_l e $f_{l'}$.

dell'altro:

$$\lambda_{12}(f) = -\frac{\frac{df_2}{d\alpha}}{\frac{df_1}{d\alpha}} \quad e \quad \lambda_{21}(f) = -\frac{\frac{df_1}{d\alpha}}{\frac{df_2}{d\alpha}} \Rightarrow \lambda_{12}(f) = \frac{1}{\lambda_{21}(f)} \text{ per ogni } f \in F$$

Vediamo ora altre due possibili espressioni del tasso marginale di sostituzione.

Tasso di sostituzione e funzione di utilità Lungo una curva di indifferenza, la funzione di utilità è costante: $u(f(\alpha)) = c$ per ogni α . Quindi, ha derivata nulla rispetto ad α :

$$\frac{du(f_1(\alpha), f_2(\alpha))}{d\alpha} = 0 \Rightarrow \frac{\partial u}{\partial f_1} \frac{df_1}{d\alpha} + \frac{\partial u}{\partial f_2} \frac{df_2}{d\alpha} = 0 \Rightarrow -\frac{\frac{df_2}{d\alpha}}{\frac{df_1}{d\alpha}} = \frac{\frac{\partial u}{\partial f_1}}{\frac{\partial u}{\partial f_2}}$$

Dall'Equazione (6.1), risulta quindi che:

$$\lambda_{12}(f) = \frac{\frac{\partial u}{\partial f_1}}{\frac{\partial u}{\partial f_2}} \quad (6.2)$$

cioè il tasso marginale di sostituzione di f_1 con f_2 è il rapporto delle derivate parziali dell'utilità rispetto a f_1 e f_2 . Se l'utilità dipende molto da f_1 e poco da f_2 , il tasso di sostituzione è alto, cioè occorre una forte variazione di f_2 per compensarne una piccola di f_1 ⁴.

Tasso di sostituzione e curve di indifferenza Se le curve di indifferenza sono invertibili, allora ad ogni valore del parametro α corrisponde un diverso valore di f_1 e di f_2 e si può definire la funzione $\alpha = \alpha(f_1)$, da cui deriva un'espressione esplicita $f_2 = f_2(\alpha(f_1))$ per la curva di indifferenza. La derivata di tale espressione è

$$\frac{df_2}{df_1} = \frac{df_2}{d\alpha} \frac{d\alpha}{df_1} = \frac{\frac{df_2}{d\alpha}}{\frac{df_1}{d\alpha}}$$

da cui

$$\lambda_{12}(f) = -\frac{df_2}{df_1} \quad (6.3)$$

cioè il tasso marginale di sostituzione è la pendenza delle curve di indifferenza cambiata di segno.

6.4.7 Additività e tasso di sostituzione

Torniamo al problema di quali relazioni di preferenza con $p = 2$ indicatori siano additive. Combinando due valori distinti per l'indicatore f_1 (rispettivamente, f'_1 e f''_1) e due valori distinti per l'indicatore f_2 (f'_2 e f''_2), si possono costruire quattro impatti diversi. In generale, i tassi marginali di sostituzione λ_{12} in questi quattro impatti sono del tutto generici. Tuttavia, può avvenire che siano legati fra loro.

⁴L'Equazione (6.2) vale anche quando ci sono $p \geq 3$ indicatori, dato che si definisce il tasso di sostituzione mantenendo tutti gli indicatori costanti tranne due.

Definizione 26 Si dice regolarità degli incrementi la proprietà per cui:

$$\lambda_{12}(f'_1, f'_2) \lambda_{12}(f''_1, f''_2) = \lambda_{12}(f''_1, f'_2) \lambda_{12}(f'_1, f''_2)$$

per ogni quaterna di impatti $(f'_1, f'_2), (f''_1, f''_2), (f''_1, f'_2), (f'_1, f''_2) \in F$.

Per qualsiasi rettangolo in F , la regolarità degli incrementi richiede che i tassi marginali nei vertici opposti abbiano lo stesso prodotto. Il nome di questa proprietà risulta ovvio se si esprime la relazione nei seguenti modi, del tutto equivalenti:

$$\frac{\lambda(f'_1, f'_2)}{\lambda(f''_1, f'_2)} = \frac{\lambda(f'_1, f''_2)}{\lambda(f''_1, f''_2)} \quad \text{e} \quad \frac{\lambda(f'_1, f'_2)}{\lambda(f'_1, f''_2)} = \frac{\lambda(f''_1, f'_2)}{\lambda(f''_1, f''_2)}$$

cioè, quando si passa da f'_1 a f''_1 , il tasso di sostituzione varia dello stesso fattore moltiplicativo qualunque sia il valore di f_2 ; analogamente, quando si passa da f'_2 a f''_2 , il tasso di sostituzione varia dello stesso fattore moltiplicativo qualunque sia f_1 .

La Figura 6.5 illustra la relazione: nei quattro punti P, Q, R e S i tassi marginali di sostituzione sono, rispettivamente, $-a/c$, $-b/c$, $-a/d$ e $-b/d$. Il prodotto dei tassi in P e S è ab/cd , e coincide col prodotto dei tassi nei punti Q e R . Infatti, passando da P a Q il tasso di sostituzione cambia di un fattore moltiplicativo pari a b/a ; passando da R a S i due tassi di sostituzione sono ancora diversi, ma legati dallo stesso fattore moltiplicativo b/a . Lo stesso succede muovendosi lungo l'asse f_2 , cioè da P a R e da Q a S .

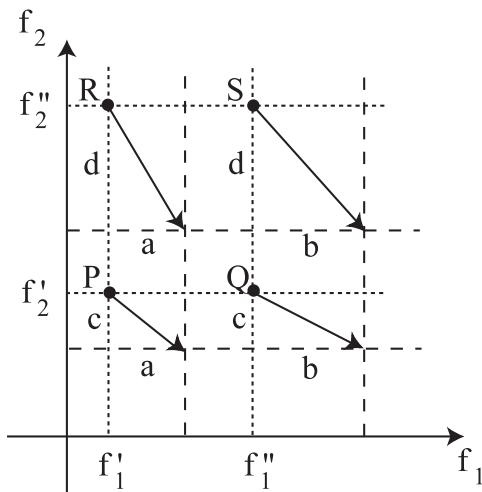


Figura 6.5: Vale la regolarità degli incrementi: il rapporto fra le variazioni di f_2 e f_1 va crescendo al crescere di f_1 in modo regolare (indipendente da f_1) e va calando al crescere di f_2 in modo regolare (indipendente da f_2)

Esempio 45 Riprendiamo la relazione di preferenza della Figura 6.4:

$$(2, 2) \prec (2, 1) \prec (1, 2) \prec (2, 0) \sim (1, 1) \sim (0, 2) \prec (1, 0) \sim (0, 1) \prec (0, 0)$$

che gode della mutua indipendenza preferenziale, ma non dell'additività. Consideriamo i quattro punti $(1, 0)$, $(2, 0)$, $(1, 1)$ e $(2, 1)$: nei primi tre il tasso di sostituzione è unitario (per bilanciare un calo unitario di f_1 occorre una crescita unitaria di f_2 ; in $(2, 1)$, invece, il tasso di sostituzione è strettamente maggiore di 1 (per bilanciare un calo unitario di f_1 non basta una crescita unitaria di f_2). Quindi non vale la regolarità degli incrementi.

Esempio 46 Consideriamo invece la relazione di preferenza le cui curve di indifferenza sono descritte in Figura 6.3 con la funzione di utilità di Cobb-Douglas $u(f_1, f_2) = f_1^2 f_2^3$. Il tasso di sostituzione è:

$$\lambda_{12}(f) = \frac{\frac{\partial u}{\partial f_1}}{\frac{\partial u}{\partial f_2}} = \frac{2f_1 f_2^3}{3f_1^2 f_2^2} = \frac{2f_2}{3f_1}$$

per cui vale la condizione di regolarità degli incrementi:

$$\lambda_{12}(f'_1, f'_2) \lambda_{12}(f''_1, f''_2) = \frac{2f'_2}{3f'_1} \frac{2f''_2}{3f''_1} = \frac{2f'_2}{3f'_1} \frac{2f''_2}{3f''_1} = \lambda_{12}(f'_1, f''_2) \lambda_{12}(f''_1, f'_2)$$

Infatti, tale relazione è esprimibile anche con una funzione di utilità additiva, come si è visto.

Ora esploriamo la relazione fra regolarità degli incrementi e additività.

Teorema 20 Se una relazione di preferenza ammette una funzione di utilità additiva, essa gode della regolarità degli incrementi rispetto a ogni coppia di indicatori.

Dimostrazione. Si tratta ancora una volta di applicare la definizione:

$$\lambda_{12}(f'_1, f'_2) \lambda_{12}(f''_1, f''_2) = \frac{\frac{\partial u}{\partial f'_1}}{\frac{\partial u}{\partial f'_2}} \frac{\frac{\partial u}{\partial f''_1}}{\frac{\partial u}{\partial f''_2}}$$

Siccome $u(f_1, f_2) = u_1(f_1) + u_2(f_2)$:

$$\lambda_{12}(f'_1, f'_2) \lambda_{12}(f''_1, f''_2) = \frac{\frac{du_1}{df'_1}}{\frac{du_1}{df'_2}} \cdot \frac{\frac{du_1}{df''_1}}{\frac{du_1}{df''_2}} = \frac{\frac{du_1}{df'_1}}{\frac{du_2}{df'_2}} \cdot \frac{\frac{du_1}{df''_1}}{\frac{du_2}{df''_2}}$$

e passando nuovamente alla funzione $u(f_1, f_2)$ si ottiene:

$$\lambda_{12}(f'_1, f'_2) \lambda_{12}(f''_1, f''_2) = \frac{\frac{\partial u}{\partial f'_1}}{\frac{\partial u}{\partial f''_2}} \frac{\frac{\partial u}{\partial f''_1}}{\frac{\partial u}{\partial f'_2}} = \lambda_{12}(f'_1, f''_2) \lambda_{12}(f''_1, f'_2)$$

■

Quindi la regolarità degli incrementi è necessaria all'additività. Ci si può ora chiedere se non sia anche sufficiente, almeno sotto opportune condizioni. La risposta è positiva, proprio nel caso che ci mancava.

Teorema 21 Se $p = 2$ e se vale la regolarità degli incrementi fra i due indicatori, esiste una funzione di utilità additiva.

Esempio 47 Le funzioni di Cobb-Douglas godono della regolarità degli incrementi nonostante non siano funzioni di utilità additive. Tuttavia, esse sono equivalenti alle funzioni additive $u'(f) = \sum_{l=1}^p \alpha_l \log f_l$, ottenute con la semplice trasformazione $u'(f) = \log u(f)$.

Tasso marginale di sostituzione uniforme

Un caso particolare notevole che gode della regolarità degli incrementi è quello in cui il tasso marginale di sostituzione non dipende da f , ma è uniforme ($\lambda_{12}(f) = \bar{\lambda}$), cioè in cui il decisore è sempre disposto a sostituire un'unità di f_1 con $\bar{\lambda}$ unità di f_2 . In tal caso, esiste la funzione di utilità additiva:

$$u(f_1, f_2) = w_1 f_1 + w_2 f_2$$

e il tasso marginale di sostituzione è $\lambda_{12} = \frac{w_1}{w_2}$. In questo caso, le curve di indifferenza sono dei fasci di rette parallele.

6.4.8 Costruzione di una funzione di utilità additiva

Supponiamo che la teoria garantisca l'esistenza di una funzione di utilità additiva $u(f) = \sum_{l=1}^p u_l(f_l)$. Tuttavia, essa non fornisce i termini $u_l(f_l)$ e non basta trovare generiche funzioni di utilità valide per ciascun attributo; bisogna anche combinarle correttamente. Il procedimento per farlo si può scomporre in due fasi:

1. costruire funzioni $\tilde{u}_l(f_l)$ che rispettino la regolarità degli incrementi, ma non specifichino la scala assoluta dei valori di utilità;
2. valutare i tassi di sostituzione, in modo da riscalare le funzioni \tilde{u}_l prima di sommarle.

In altre parole, prima si determinano delle funzioni $\tilde{u}_l(f_l)$ che assumono valori relativi a ciascun indicatore; poi si determinano dei coefficienti w_l per riscalarle prima di sommarle:

$$u(f) = \sum_{l=1}^p w_l \tilde{u}_l(f_l)$$

dove i coefficienti w_l danno luogo a una combinazione convessa: $w_l \geq 0$ per ogni $l \in P$ e $\sum_{l=1}^p w_l = 1$. In genere, le funzioni $\tilde{u}_l(f_l)$ sono *normalizzate*, cioè assumono valori che:

1. sono numeri puri, adimensionali;
2. cadono nell'intervallo $[0, 1]$.

Il vantaggio principale di tali funzioni è che rimuovono la *scala*, cioè il valore di origine (*offset*) e l'unità della misura.

Metodo della frazione media

Costruire una funzione utilità per un impatto monodimensionale ($F \subseteq \mathbb{R}$) è molto più semplice che farlo per un impatto multidimensionale. Si tratta infatti solo di confrontare coppie di numeri e chiedersi quale dei due sia migliore: può essere noioso, ma non è impossibile rispondere in maniera coerente a tutte le domande necessarie.

Tuttavia la funzione che vogliamo costruire deve rispettare una condizione più forte della semplice richiesta di ordinare gli impatti. Vogliamo che la differenza fra le utilità di impatti diversi misuri la forza della relativa preferenza: impatti quasi indifferenti devono avere valori di utilità molto simili anche se il valore degli indicatori è molto diverso; viceversa, impatti preferenzialmente ben distinti devono avere valori di utilità molto diversi anche se il valore degli indicatori è simile.

Il metodo della frazione media costruisce una simile funzione in modo dicotomico, per dimezzamenti successivi. Supponiamo di dover regolare il condizionamento di una stanza, e concentriamoci in particolare sulla temperatura. Essa non è né un costo né un beneficio, dato che si preferisce un valore né troppo alto né troppo basso. Anzi tutto, occorre determinare la proiezione sull'asse f_1 (temperatura) dell'insieme degli impatti F fissando i valori estremi (per esempio, fra $f = 16$ e $f = 30$ gradi). Poi si domanda al decisore quali temperature corrispondono all'utilità minima e massima. Alle temperature (che possono essere più d'una) pessime si associa convenzionalmente un valore di utilità pari a 0, mentre a quelle ottime un valore pari a 1. Per esempio, sia $u(30) = 0$ e $u(22) = 1$.

A questo punto, si domanda al decisore di indicare le temperature che abbiano utilità esattamente intermedia fra quelle estreme, e si assegna loro un valore di utilità pari a 0.5. Supponiamo che egli indichi $u(27) = u(18) = 0.5$. Si procede dimezzando via via gli intervalli di utilità finché si è ottenuto un numero di punti sufficiente a tentare l'interpolazione con un'espressione analitica $\tilde{u}(f)$ tratta da un'opportuna famiglia di funzioni.

Utilità proporzionale all'indicatore

Nel caso di indicatori che rappresentino costi o benefici, a volte si può ipotizzare che la preferenza sia approssimativamente proporzionale al valore dell'indicatore stesso. Un modo comune di generare una funzione di utilità normalizzata $\tilde{u} : F \rightarrow \mathbb{R}$ allora è il seguente:

$$\hat{u}(f) = \frac{f - \min_{x \in X} f(x)}{\max_{x \in X} f(x) - \min_{x \in X} f(x)}$$

Esempio 48 Consideriamo il problema del viaggio con due indicatori (tempo e costo) discusso nell'Esempio ?? del Capitolo 5. Ipotizziamo che l'utilità sia additiva e abbia una dipendenza lineare sia dal tempo sia dal costo.

	f_1 (ore)	f_2 (Euro)	\tilde{u}_1	\tilde{u}_2
Treno	5.5	100	0.00	1.00
Auto	4.0	150	0.33	0.83
Aereo	1.0	300	1.00	0.33
Pullman	5.0	180	0.11	0.73
Taxi	4.0	400	0.33	0.00

Tabella 6.1: Tempi e costi associati a diversi mezzi di trasporto per compiere un viaggio e loro utilità normalizzate sotto l'ipotesi di utilità additiva e lineare

La Tabella 6.1 riporta i valori originali degli indicatori e le corrispondenti utilità normalizzate per le cinque alternative del problema. Per costruire la funzione di utilità complessiva del problema restano da determinare i coefficienti w_i con i quali combinare le utilità normalizzate.

Problemi legati alla normalizzazione

Se l'insieme delle soluzioni è infinito o combinatorico, spesso non è banale conoscere l'impatto pessimo e quello ottimo, dato che si tratta di risolvere dei problemi di Programmazione Matematica. Per gestire questa difficoltà si possono sostituire i valori estremi con una stima per difetto del minimo e una stima per eccesso del massimo. Questo produce ancora una funzione di utilità normalizzata, diversa da

quella sopra riportata in quanto non copre l'intero intervallo $[0; 1]$. Se le stime sono deboli, i valori di questa utilità normalizzata risultano "compressi" in un intervallo ristretto, il che può creare problemi nel momento in cui si volesse combinare la funzione con quelle relative ad altri indicatori.

A rigore, potrebbero anche non esistere un valore pessimo e un valore ottimo: la situazione potrebbe degradarsi o migliorare indefinitamente (che non vuol dire illimitatamente: potrebbe convergere a un valore limite senza mai raggiungerlo). In questi casi, si possono fissare dei limiti convenzionali, oltre i quali qualsiasi successivo peggioramento o miglioramento viene trascurato. Per esempio, nel caso delle concentrazioni di agenti inquinanti, esisterà un valore minimo sotto il quale si suppone (in base a indagini storiche, probabilmente) che la concentrazione sia innocua e un valore massimo sopra il quale la concentrazione è dichiarata del tutto inammissibile. La scelta di questi valori, come vedremo, può influire sul risultato del processo decisionale.

Si noti anche che l'utilità normalizzata dipende intrinsecamente dalla definizione dell'insieme degli impatti F . Questa può sembrare un'ovvia innocua, dato che F è un dato. Tuttavia, sappiamo che nei problemi decisionali complessi spesso le alternative non sono tutte note a priori. Nuove alternative, e quindi nuovi impatti, possono presentarsi in fasi successive del processo di decisione. Se i valori degli indicatori eccedono il massimo o il minimo su F , l'utilità normalizzata va ricalcolata e questo può influire sulle elaborazioni successive.

Linearizzazione delle curve di indifferenza

Se le funzioni di utilità $\tilde{u}_l(f_l)$ sono state stimate correttamente, la funzione complessiva $u(f)$ ha una caratteristica particolarmente utile: le curve di indifferenza nello spazio delle funzioni di utilità \tilde{u}_l diventano lineari, anche se le curve originali nello spazio degli indicatori f_l non lo erano. Questo perché le trasformazioni $\tilde{u}_l = \tilde{u}_l(f_l)$ distorcono opportunamente la scala sugli assi in modo da racchiudere in sé e far sparire tutta la non linearità.

Quindi si possono definire dei tassi marginali di sostituzione $\tilde{\lambda}_{lm}$ fra le componenti \tilde{u}_l della funzione di utilità, anziché fra gli indicatori f_l . Questi tassi di sostituzione sono uniformi e ciascuno coincide col rapporto fra i pesi delle due componenti:

$$\tilde{\lambda}_{lm} = \frac{\frac{\partial u}{\partial u_l}}{\frac{\partial u}{\partial u_m}} = \frac{w_l}{w_m}$$

Questa cosa semplifica moltissimo la fase di determinazione dei coefficienti w_l con cui combinare le funzioni di utilità normalizzate. Infatti, ogni coppia di impatti indifferenti determina un'equazione lineare sui coefficienti stessi.

Determinazione dei pesi

Determinate le singole componenti di una funzione di utilità additiva, restano da determinare i pesi w_l con cui combinarle. A tale scopo, occorre conoscere un numero sufficiente di coppie di impatti indifferenti. Uguagliando le utilità dei due impatti, si impongono vincoli sui pesi w_l . Quando il sistema di vincoli ammette una sola soluzione, essa fornisce il valore dei pesi. Questo procedimento è identico a quello descritto nella Sezione 6.4.2 per le funzioni di utilità generiche, ma c'è una grossa differenza: il sistema di equazioni da risolvere è lineare, anziché generico e dipendente dalla famiglia di funzioni di utilità scelta. Inoltre, i coefficienti w_l hanno un significato molto più intuitivo dei parametri α_r richiesti dal metodo generale.

Il modo più semplice di procedere è quello di concentrarsi su due attributi f_l e f_m e individuare due impatti indifferenti che differiscono solo per il valore di tali attributi. L'equazione che ne risulta determina univocamente il tasso marginale di sostituzione $\tilde{\lambda}_{lm} = w_l/w_m$. In $p-1$ passi si determinano tutti i tassi di sostituzione, e quindi tutti i pesi (a meno di una costante moltiplicativa che si determina con la condizione di normalizzazione $\sum_l w_l = 1$).

Ovviamente, non bisogna considerare coppie (l, m) per le quali il tasso di sostituzione è già noto, direttamente o indirettamente. Ad esempio, noti $\tilde{\lambda}_{12} = w_1/w_2$ e $\tilde{\lambda}_{23} = w_2/w_3$, non occorre determinare $\tilde{\lambda}_{13} = w_1/w_3$. Per evitarlo, basta costruire un grafo ausiliario, in cui i vertici rappresentano gli attributi e i lati i tassi di sostituzione marginali. Per determinare tutti i pesi, basta conoscere un sottoinsieme di $p-1$ tassi di sostituzione che formino un albero ricoprente. Infatti, ogni ciclo corrisponde a un insieme di tassi di sostituzione il cui prodotto è unitario, per cui ciascuno coincide con il reciproco del prodotto degli altri.

Se non si conoscono coppie di impatti indifferenti, si possono confrontare coppie generiche di impatti, che corrispondono a disuguaglianze sui tassi, ottenendo così intervalli entro i quali i tassi devono cadere, e stringerli via via. Non sempre è grave che i pesi siano incerti. Risolto il problema con i pesi stimati, conviene sempre compiere un'analisi di sensitività della soluzione ottima, cioè rappresentare il problema nello spazio dei pesi e determinare il supporto della soluzione ottima (l'insieme di pesi per i quali essa rimane ottima). Se i pesi stimati cadono ben dentro il supporto, pesi non molto diversi porterebbero alla stessa soluzione. In caso contrario, l'analisi di sensitività consente spesso di concentrarsi su un piccolo numero di soluzioni.

Esempio 49 Si supponga che interviste con il decisore abbiano stabilito la validità dell'ipotesi di mutua indipendenza preferenziale per un problema con quattro attributi e che si siano determinate con il metodo della frazione media le seguenti funzioni di utilità normalizzate per i singoli attributi:

$$\begin{aligned}\tilde{u}_1 &= 2f_1 + \frac{1}{2}, \quad \text{con } f_1 \in \left(-\frac{1}{4}; \frac{1}{4}\right) & \tilde{u}_2 &= \frac{\log_2 f_2 - 2}{8}, \quad \text{con } f_2 \in (4; 1024) \\ \tilde{u}_3 &= \frac{20 - f_3}{15}, \quad \text{con } f_3 \in (5; 20) & \tilde{u}_4 &= \frac{\sqrt{f_4}}{6}, \quad \text{con } f_4 \in (0; 36)\end{aligned}$$

Infine, si supponga che si siano individuate le seguenti $p-1 = 3$ coppie di impatti indifferenti:

$$\begin{aligned}A &= (0, 8, 10, 9) \sim B = \left(\frac{1}{4}, 8, 10, 4\right) \\ C &= \left(-\frac{1}{4}, 256, 10, 0\right) \sim D = \left(-\frac{1}{4}, 256, 20, 16\right) \\ E &= (0, 4, 5, 25) \sim F = \left(-\frac{1}{4}, 64, 5, 25\right)\end{aligned}$$

Si noti che si sono scelte coppie di impatti che differiscono solo per due dei quattro indicatori: A e B differiscono per il primo e quarto indicatore, C e D per il terzo e quarto, E e F per i primi due. Questo non è rigorosamente necessario, ma facilita i calcoli.

Anzi tutto, dobbiamo convertire gli impatti nelle corrispondenti funzioni di utilità.

$$\tilde{u}_A = \left(\frac{1}{2}, \frac{1}{8}, \frac{2}{3}, \frac{1}{2}\right) \quad \tilde{u}_B = \left(1, \frac{1}{8}, \frac{2}{3}, \frac{1}{3}\right)$$

$$\begin{aligned}\tilde{u}_C &= \left(0, \frac{3}{4}, \frac{2}{3}, 0\right) & \tilde{u}_D &= \left(0, \frac{3}{4}, 0, \frac{2}{3}\right) \\ \tilde{u}_E &= \left(\frac{1}{2}, 0, 1, \frac{5}{6}\right) & \tilde{u}_F &= \left(0, \frac{1}{2}, 1, \frac{5}{6}\right)\end{aligned}$$

L'additività consente di esprimere l'utilità complessiva come combinazione convessa delle singole componenti:

$$u = w_1 \tilde{u}_1 + w_2 \tilde{u}_2 + w_3 \tilde{u}_3 + w_4 \tilde{u}_4$$

da cui le tre uguaglianze

$$\begin{cases} u(A) = u(B) \Rightarrow \frac{1}{2}w_1 + \frac{1}{8}w_2 + \frac{2}{3}w_3 + \frac{1}{2}w_4 = 1w_1 + \frac{1}{8}w_2 + \frac{2}{3}w_3 + \frac{1}{3}w_4 \\ u(C) = u(D) \Rightarrow 0w_1 + \frac{3}{4}w_2 + \frac{2}{3}w_3 + 0w_4 = 0w_1 + \frac{3}{4}w_2 + 0w_3 + \frac{2}{3}w_4 \quad \Rightarrow \\ u(E) = u(F) \Rightarrow \frac{1}{2}w_1 + 0w_2 + 1w_3 + \frac{5}{6}w_4 = 0w_1 + \frac{1}{2}w_2 + 1w_3 + \frac{5}{6}w_4 \end{cases}$$

$$\Rightarrow \begin{cases} \frac{1}{6}w_4 = \frac{1}{2}w_1 \\ \frac{2}{3}w_3 = \frac{2}{3}w_4 \\ \frac{1}{2}w_1 = \frac{1}{2}w_2 \end{cases} \Rightarrow \begin{cases} w_4 = 3w_1 \\ w_3 = w_4 \\ w_1 = w_2 \end{cases}$$

Avendo scelto gli impatti in modo che solo due indicatori differissero, le uguaglianze legano due pesi alla volta. In questo modo, ne determinano i tassi di sostituzione $\hat{\lambda}_{lm} = w_l/w_m$. Per esempio, $\hat{\lambda}_{41} = 3$. Si noti come il peso w_4 sia superiore al peso w_1 , suggerendo che l'utilità \tilde{u}_4 sia più rilevante dell'utilità \tilde{u}_1 . Infatti, passando dall'impatto A all'impatto B, l'utilità \tilde{u}_4 varia poco rispetto all'utilità \tilde{u}_1 . Se considerassimo direttamente gli indicatori f_4 e f_1 , vedremmo esattamente l'opposto (f_4 cambia molto rispetto a f_1), ma gli indicatori dipendono dalle unità di misura e influenzano l'utilità in modo diverso in impatti diversi.

Le uguaglianze determinate dalle coppie di impatti indifferenti permettono di calcolare i pesi, aggiungendo la condizione di normalizzazione:

$$w_1 + w_2 + w_3 + w_4 = 1 \Rightarrow w_2 + w_2 + 3w_2 + 3w_2 = 1 \Rightarrow w_2 = \frac{1}{8}$$

da cui

$$w = \left[\frac{1}{8} \frac{1}{8} \frac{3}{8} \frac{3}{8} \right]^T$$

Relazioni fra teoria dell'utilità a molti attributi e metodo dei pesi

Ci sono forti analogie fra questo procedimento e il metodo dei pesi per determinare le soluzioni paretiane, ma il contesto è del tutto diverso:

1. il metodo dei pesi non richiede preferenze complete, né mutua indipendenza preferenziale o additività degli incrementi, ma solo che gli attributi siano funzioni di utilità o di costo;
2. il metodo dei pesi non richiede di determinare un vettore di pesi w , ma scorre tutti i vettori possibili;

3. il metodo dei pesi non fornisce una soluzione “ottima”, ma un sottoinsieme della regione paretiana;
4. il metodo dei pesi combina direttamente gli attributi, senza filtrarli attraverso funzioni di utilità opportunamente costruite per riflettere i le utilità relative associate a valori diversi di un indicatore; di conseguenza, gli risultano inaccessibili le soluzioni non supportate; al contrario, opportune funzioni di utilità, e quindi curve di indifferenza di forma opportuna, consentono alla funzione di utilità globale di ammettere tali soluzioni come ottime.

Esempio 50 Si consideri il problema di albero ricoprente minimo discusso nell’Esempio 28 della Sezione 5.6, e si supponga che i due attributi, che rappresentano entrambi dei costi, siano associati alle seguenti funzioni di utilità

$$u_1(f_1) = \sqrt{9 - f_1} \quad u_2(f_2) = \sqrt{23 - f_3}$$

Il problema ammette solo tre soluzioni ammissibili, che corrispondono agli impatti $f_A = (4, 23)$, $f_B = (7, 21)$ e $f_C = (9, 18)$.

FIGURA DA AGGIUNGERE, CON LE CURVE DI INDIFFERENZA NELLO SPAZIO DEGLI INDICATORI E NELLO SPAZIO DELLE UTILITÀ

Tutte e tre le soluzioni sono paretiane, ma il metodo dei pesi è in grado di identificare solo A e C, perché B è una soluzione non supportata, annidata in una concavità dell’insieme degli impatti, e non è ottima per alcuna c combinazione lineare dei due costi. Al contrario, curve di indifferenza non lineari possono far sì che $f(B)$ sia in realtà preferibile a $f(A)$ e $f(C)$. Corrispondentemente, una funzione di utilità non lineare come quella sopra menzionata permette di distorcere le curve di indifferenza nello spazio delle utilità, rendendo esplicito il fatto che anche B è paretiana, e che può essere ottima per opportune funzioni di utilità. Infatti,

$$u(f) = u_1(f_1) + u_2(f_2) \Rightarrow \begin{cases} u(f(A)) = \sqrt{5} \\ u(f(B)) = 2\sqrt{2} \\ u(f(C)) = \sqrt{5} \end{cases}$$

so that $u(f(B)) > u(f(A)) = u(f(C))$, and $f(B) \prec f(A) \sim f(C)$.

In linea di principio, ogni soluzione paretiana ammette una funzione di utilità per la quale essa è ottima.

6.4.9 Il procedimento in sintesi

Per riassumere, il procedimento di determinazione della funzione di utilità si articola nei seguenti passi:

1. si intervista il decisore per capire se vale la mutua indipendenza preferenziale (chiedendogli di confrontare diverse coppie di indicatori con il sottoinsieme complementare);
2. in caso positivo, ci si concentra su ciascun attributo, fissando gli altri a un valore qualsiasi, e si costruisce una componente monodimensionale della funzione utilità con un metodo che rispetti l’entità dei divari di preferenza;
3. si determinano i pesi con i quali combinare tali funzioni, campionando gli impatti fino a trovare un numero sufficiente di coppie di impatti indifferenti;
4. si verifica a posteriori che la funzione utilità ottenuta sia valida testandola su impatti non utilizzati in precedenza.

In qualsiasi momento, può essere necessario tornare indietro a correggere le singole funzioni componenti, oppure addirittura falsificare l'ipotesi di additività, nel qual caso l'intero procedimento viene invalidato.

6.5 Esercizi* †

Esercizio 1

Si consideri il seguente problema con due indicatori che rappresentano costi:

$$\begin{aligned} \min f_1 &= \frac{1}{4}(x_1 - 4)^2 + \frac{1}{4}x_2^2 \\ \min f_2 &= 2 - x_2 \\ 2x_1 + x_2 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

Si determini la soluzione ottima rispetto a una funzione di utilità con tasso marginale di sostituzione uniforme fra f_1 e f_2 e pesi $w_1 = 1/3$ e $w_2 = 2/3$.

Soluzione

La soluzione ottima è $x^* = (4/5, 12/5)$ e vale $f^* = (4, -2/5)$.

Esercizio 2

Si consideri il seguente problema con due indicatori che rappresentano benefici:

$$\begin{aligned} \max f_1 &= x_1 - x_2 \\ \max f_2 &= x_2 \\ x_1 + x_2 &\leq 3 \\ 0 \leq x_1 &\leq 2 \\ x_2 &\geq 0 \end{aligned}$$

Si determini la soluzione ottima rispetto alla funzione di utilità che combina i due indicatori con pesi $w_1 = 1/3$ e $w_2 = 2/3$.

Soluzione

La soluzione ottima è $x^* = (2, 1)$ e vale $f^* = (1, 1)$.

Esercizio 3

Si consideri il problema:

$$\begin{aligned} \max f_1 &= x_1 + 3x_2 \\ \max f_2 &= -3x_1 - 2x_2 \\ 2x_1 + x_2 &\leq 32 \\ x_1 + x_2 &\leq 20 \\ x_1 + 5x_2 &\leq 72 \\ x_1, x_2 &\geq 0 \end{aligned}$$

Si determini la soluzione ottima rispetto alla funzione di utilità $u(f_1, f_2) = 2f_1 + f_2$.

*Le soluzioni di questi esercizi non sono state ancora riviste: eventuali segnalazioni di errore sono benvenute.

†Devo molti di questi esercizi a temi d'esame del prof. Alberto Colomni.

Soluzione

La soluzione ottima è $x^* = (0, 72/5)$ e vale $f^* = (216/5, -144/5)$.

Esercizio 4

Si determini la soluzione preferita assegnando al primo indicatore peso pari alla metà del secondo ($w_1 = w_2/2$).

$$\begin{aligned} \max f_1 &= -x_1 - x_2 \\ \max f_2 &= x_1 \\ 3x_1^2 + 4x_2 &\leq 12 \\ x_2 &\geq 0 \end{aligned}$$

Soluzione

La soluzione ottima è $x^* = (2, 0)$ e vale $f^* = (-2, 2)$.

Esercizio 5

Si risolva graficamente il seguente problema di decisione nello spazio delle variabili:

$$\begin{aligned} \max f_1 &= 9x_1^2 + 4x_2^2 - 18x_1 - 16x_2 \\ \max f_2 &= -x_1 \\ 3x_1 + x_2 &\leq 6 \\ 3x_1 + 2x_2 &\leq 9 \\ x_2 &\geq 0 \end{aligned}$$

rispetto alla funzione di utilità $u(f) = f_1 - 9f_2^2$.

Soluzione

La soluzione ottima è $x^* = (19/12, 15/12)$ con $u^* = -169/4$.

Esercizio 6

La tabella seguente rappresenta le prestazioni di cinque alternative rispetto a quattro criteri decisionali (tutti da massimizzare), in una scala di valori tra 0 e 100.

	a_1	a_2	a_3	a_4	a_5
f_1	100	70	60	40	20
f_2	60	45	40	100	80
f_3	60	25	20	80	100
f_4	20	100	90	50	40

Quale alternativa risulta la migliore se le curve di indifferenza sono del tipo:

$$u(f) = w_1 f_1 + w_2 f_2 + w_3 f_3 + w_4 f_4 \text{ con } w_i = 0.25 \text{ per } i = 1, \dots, 4$$

Di quanto bisogna aumentare il valore di w_1 (mantenendo costanti i valori degli altri coefficienti) affinché a_1 risulti l'alternativa migliore? E il valore di w_4 ?

Soluzione

L'alternativa ottima è a_4 , con $u(f(a_4)) = 67.5$, mentre per le altre alternative risulta $u(f(a_i)) = 60$ per ogni i .

A rigore, ogni variazione di un coefficiente di peso dovrebbe essere bilanciata da una variazione opposta degli altri, in modo da conservare la condizione di normalizzazione. Per semplicità, questo esercizio ignora il vincolo⁵. Per rendere a_1 l'alternativa migliore, bisogna alzare w_1 da 0.25 e 0.375, mentre alzare w_4 non può mai rendere a_1 l'alternativa migliore, dato che essa è peggiore delle altre alternative rispetto al quarto criterio.

Esercizio 7

Si consideri il problema:

$$\begin{aligned} \max f_1(x) &= -x_1 + 2x_2 \\ \max f_2(x) &= 2x_1 - x_2 \\ x_1 + x_2 &\leq 7 \\ -x_1 + x_2 &\leq 3 \\ x_1 - x_2 &\leq 3 \\ 0 \leq x_1 &\leq 4 \\ 0 \leq x_2 &\leq 4 \end{aligned}$$

e si determini la soluzione ottima rispetto alla funzione di utilità $u(f) = w_1f_1 + w_2f_2$ con $w_1 = 0.25$ e $w_2 = 0.75$.

Si supponga poi di aggiungere un terzo obiettivo, anch'esso da massimizzare, $f_3(x) = 2x_1 + x_2$ e si disegnino nel piano w_1w_2 le regioni nelle quali ciascuna soluzione di base ammissibile del problema di programmazione lineare risulta ottima (ce ne sono sette)⁶.

Soluzione

La soluzione ottima è $x^* = (4, 1)$.

Poiché $w_3 = 1 - w_1 - w_2$, si può rappresentare la regione W nel piano w_1w_2 come $W = \{w \in \mathbb{R}^2 : w_1 + w_2 \leq 1, w_1 \geq 0, w_2 \geq 0\}$. Vi sono quattro supporti:

1. per $w_1 \geq 2/3$, la soluzione ottima è $(1, 4)$;
2. per $w_2 \leq 2w_1 - 1/2$ e $w_1 \leq 2/3$, la soluzione ottima è $(3, 4)$;
3. per $w_2 \leq 1/2w_1 + 1/2$ e $w_2 \geq 2w_1 - 1/2$, la soluzione ottima è $(4, 3)$;
4. per $w_2 \geq 1/2w_1 + 1/2$, la soluzione ottima è $(4, 1)$.

⁵Si provi a risolverlo anche rispettando la condizione di normalizzazione: all'esame probabilmente verrà richiesto di farlo. Ad ogni modo, finché la scelta è esplicita, entrambi gli approcci sono corretti.

⁶Questo esercizio si può risolvere graficamente oppure con il metodo del simplesso, se lo si conosce.

Esercizio 8

Dato il seguente problema di decisione:

$$\begin{aligned} \max f_1(x) &= x_1 - 3x_2 \\ \max f_2(x) &= -4x_1 + x_2 \\ -2x_1 + 2x_2 &\leq 7 \\ 2x_1 + 2x_2 &\leq 11 \\ x_1 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

si determini la soluzione ottima supponendo che il decisore dichiari un tasso marginale di sostituzione uniforme di 4 unità di f_1 per 1 unità di f_2 .

Soluzione

Con $\lambda_{12} = 1/4$ è $u(f) = f_1 + 4f_2$. La soluzione ottima è $x^* = (0, 7/2)$.

Esercizio 9

È dato il problema:

$$\begin{aligned} \min f_1(x) &= x_1^2 + x_2^2 \\ \max f_2(x) &= x_2 \\ x_2 &\leq 10 \end{aligned}$$

Si esprima analiticamente l'utilità associata a $f_1(x)$ come:

$$u_1(f_1) = \begin{cases} 10 - \frac{f_1}{20} & \text{per } 0 \leq f_1 \leq 200 \\ 0 & \text{per } f_1 \geq 200 \end{cases}$$

cioè tale utilità decresce linearmente fino ad annullarsi per $f_1 = 200$, mentre l'utilità associata a f_2 coincide con il valore stesso di f_2 .

Si determini la soluzione ottima rispetto a questa funzione di utilità.

Si indichino le coordinate del “punto utopia” nello spazio delle utilità e nello spazio degli indicatori.

Soluzione

Le componenti della funzione di utilità sono

$$u_1(x) = \max \left[10 - \frac{x_1^2 + x_2^2}{20}, 0 \right] \text{ e } u_2(x) = x_2$$

e quindi la soluzione ottima è ...

Il punto utopia nello spazio delle utilità ha come coordinate i valori massimi delle due componenti dell'utilità, cioè $(10, 10)$. Nello spazio degli indicatori, occorre riportare i valori di f_1 e f_2 che danno luogo a tali utilità, cioè $(0, 10)$.

Esercizio 10

Un centro sociale cerca una nuova sede: esistono quattro alternative (A, B, C e D) oltre all'alternativa 0 (restare nella sede attuale). Si è stabilito che la scelta tra le cinque alternative debba essere definitiva e che sarà fatta in base a tre fattori: costi, accessibilità e prestigio. È fornita una tabella indicante le utilità per ciascuna alternativa e fattore in una scala tra 0 e 100. È fornito anche un vettore di pesi dei tre fattori.

Indicatori	A	B	C	D	0		pesi
Costi	90	90	90	1	100	w_1	$1/3$
Accessibilità	12	13	10	100	37	w_2	$1/3$
Prestigio	30	1	5	100	10	w_3	$1/3$

Si ordinino le alternative in base alle informazioni date.

Si faccia un'analisi di sensitività rispetto al peso dei costi, per stabilire l'intervallo di pesi in cui l'alternativa scelta non cambia.

Si rappresentino nello spazio dei pesi i supporti delle diverse alternative.

Soluzione

SOLUZIONE NON DISPONIBILE

Esercizio 11

È dato il problema:

$$\begin{aligned} \min f_1(x) &= x^2 - 4x \\ \min f_2(x) &= -x^2 \\ 0 \leq x &\leq 3 \end{aligned}$$

Si rappresenti il punto utopia U nello spazio degli obiettivi e si dica quale fra gli impatti $A' = (-4, -4)$ e $B' = (-3, -9)$ è preferibile in base alla distanza euclidea da U .⁷

Aggiungendo un terzo obiettivo $f_3 = x$, sempre da minimizzare, si dica entro quali valori del relativo peso w_3 il punto $\bar{x} = 3$ costituisce la soluzione ottima, supponendo che gli altri due pesi w_1 e w_2 siano uguali.

Soluzione

Il punto utopia è $U = (-4, -9)$. L'impatto B' è il più vicino ad U , e quindi è preferibile.

La soluzione ottima è $\bar{x} = 3$ per $w_3 \leq 16/10^8$.

⁷Il metodo del punto utopia richiederebbe di determinare la soluzione paretiana più vicina al punto utopia, ma non ho verificato quanto sia difficile questo esercizio.

⁸Ancora una volta, questo risultato ipotizza $w_1 = w_2 = 0.4$ senza rinormalizzazioni al variare di w_3 . Rinormalizzando, si otterrà un risultato diverso.

Esercizio 12

Dato il seguente problema di decisione:

$$\begin{aligned}\min f_1(x) &= x_1^2 + 4x_2^2 - 2x_1 - 16x_2 \\ \min f_2(x) &= -5x_1 - x_2 \\ x_1^2 + x_2^2 &\leq 8 \\ x_2 &\geq 0\end{aligned}$$

si indichino le coordinate del punto utopia U nello spazio degli obiettivi.

Soluzione

Le coordinate del punto U nello spazio degli obiettivi sono date rispettivamente da $\min f_1$ e $\min f_2$, cioè $U = (-17, -52/\sqrt{13})$.

Esercizio 13

È dato il seguente problema decisionale con tre alternative e quattro attributi (utilità), cui sono associati i pesi w_i ($i = 1, \dots, 4$):

Attributi	A	B	C		pesi
u_1	0	100	80	w_1	0.25
u_2	100	83	0	w_2	0.30
u_3	70	20	100	w_3	0.40
u_4	40	100	20	w_4	0.05

Si determini l'alternativa migliore rispetto alla funzione di utilità che combina gli indicatori con i pesi indicati, e si conduca un'analisi di sensitività rispetto a $w_2 = 0.30$.

Soluzione

Le alternative rispettano il seguente ordinamento: $B \prec C \prec A$.

L'alternativa B rimane la migliore per $23/80 \leq w_2 \leq 2/5$ ⁹.

Esercizio 14

Si consideri il problema:

$$\begin{aligned}\min f_1(x) &= x_1^2 + x_2^2 - 2x_1 \\ \min f_2(x) &= -x_2 \\ x_1^2 + 4x_2^2 &\leq 8 \\ x_1 - 2x_2 &\geq 0\end{aligned}$$

Si determini il punto utopia U e la soluzione finale che si sceglierrebbe secondo tale logica in base alla distanza di Manhattan L_1 ($d_{xy} = \sum_{i=1}^n |x_i - y_i|$) fra i seguenti tre impatti alternativi: $f(A) = (-1, 0)$, $f(B) = (-3/4, -1/2)$ e $f(C) = (1, -1)$.

⁹Questo risultato si ottiene probabilmente non rinormalizzando gli altri pesi al variare di w_2 .

Soluzione

Il punto utopia è $U = (-1, -1)$. L'impatto B' è il più vicino, e quindi quello da preferire in base a questo criterio di scelta.

Esercizio 15

Data la seguente matrice di valutazione e il seguente vettore di pesi

Attributi	A	B	C	pesi
u_1	80	20	60	
u_2	10	90	50	$w_1 \quad 0.7$ $w_2 \quad 0.3$

si determini l'ordinamento delle alternative e si conduca un'analisi di sensitività al variare di w_1 .

Soluzione

L'ordinamento delle alternative in base ai pesi dati è $A \prec C \prec B$.

L'alternativa A rimane ottima finché $w_1 \geq 0.6^{10}$.

¹⁰Questo risultato si ottiene probabilmente non rinormalizzando gli altri pesi al variare di w_1 .

Capitolo 7

Metodi a razionalità debole

0

7.1 Decisori parzialmente incoerenti

Se si applica tutto l'armamentario della teoria dell'utilità a molti attributi, eseguendo ogni passaggio con la dovuta precisione, nel caso additivo si ottengono p funzioni di utilità normalizzate \tilde{u}_l e p pesi w_l che ne modulano l'importanza nella funzione complessiva. I tassi di sostituzione fra le utilità normalizzate risultano uniformi e pari ai rapporti dei pesi stessi ($\tilde{\lambda}_{lm} = w_l/w_m$), mentre le curve di indifferenza nello spazio delle funzioni di utilità formano fasci di iperpiani paralleli.

In pratica, il problema è che il decisore non è generalmente in grado di stimare correttamente i tassi di sostituzione come richiesto dalla teoria, perché considera indifferenti anche coppie di impatti fra i quali esiste invece una debole preferenza. Il risultato è che il calcolo dei pesi w_l a partire da $p-1$ coppie di impatti indifferenti non è affidabile.

Supponiamo di fare qualche confronto in più per aumentare la sicurezza del risultato. Al limite, potremmo trovare per ogni coppia di indicatori (l, m) due impatti fra loro indifferenti che abbiano valori diversi per f_l e f_m e uguali per tutti gli altri indicatori. Ognuna di queste coppie fornisce un valore ipotetico per il tasso di sostituzione $\tilde{\lambda}_{lm} = w_l/w_m$.

Definizione 27 Definiamo matrice dei confronti a coppie $\tilde{\Lambda} = \{\tilde{\lambda}_{lm}\}$ la matrice che contiene le stime dei tassi di sostituzione fra le utilità normalizzate.

In presenza di un decisore perfettamente coerente, tale matrice è formata dai rapporti w_l/w_m dei pesi delle utilità normalizzate, e quindi gode delle seguenti proprietà:

1. *positività*: i tassi di sostituzione fra utilità normalizzate sono sempre positivi

$$\tilde{\lambda}_{lm} > 0 \quad \text{per ogni } l, m \in P$$

2. *reciprocità*: il tasso di sostituzione di una utilità normalizzata rispetto a un'altra è il reciproco del tasso della seconda rispetto alla prima

$$\tilde{\lambda}_{lm} = \frac{1}{\tilde{\lambda}_{ml}} \quad \text{per ogni } l, m \in P$$

3. *coerenza*: due tassi di sostituzione con indici “in sequenza” determinano il terzo

$$\tilde{\lambda}_{ln} = \tilde{\lambda}_{lm}\tilde{\lambda}_{mn} \quad \text{per ogni } l, m, n \in P$$

Da queste proprietà deriva una serie di conseguenze forti.

Proposizione 1 *Gli elementi della diagonale principale di $\tilde{\Lambda}$ sono tutti unitari.*

Dimostrazione.

$$\tilde{\lambda}_{ll} = \tilde{\lambda}_{lm}\tilde{\lambda}_{ml} = 1 \quad \text{per ogni } l \in P$$

■

Proposizione 2 *Le colonne di $\tilde{\Lambda}$ sono fra loro proporzionali, e così pure le righe.*

Dimostrazione. Il rapporto fra elementi corrispondenti delle colonne m e n è $\tilde{\lambda}_{mn}$, uniforme. Infatti, fissata qualsiasi coppia di indici m e n :

$$\frac{\tilde{\lambda}_{ln}}{\tilde{\lambda}_{lm}} = \tilde{\lambda}_{mn} \quad \text{per ogni } l \in P$$

La dimostrazione per le righe è del tutto analoga. ■

Proposizione 3 *La matrice $\tilde{\Lambda}$ ha rango 1:*

- $p - 1$ autovalori sono nulli;
- l'autovalore dominante μ^{\max} è pari a p ;
- l'autovettore dominante x^{\max} è proporzionale al vettore dei pesi.

Dimostrazione. Siccome le colonne sono tutte proporzionali, il rango della matrice è 1. Inoltre, esso coincide con il numero di autovalori non nulli, per cui quelli nulli sono $p - 1$. L'autovalore dominante μ^{\max} coincide quindi con la somma di tutti gli autovalori, che a sua volta coincide con la *traccia* della matrice, ovvero la somma degli elementi della diagonale principale. Questi sono tutti unitari e il loro numero è pari a p . Infine, l'autovettore dominante x^{\max} risolve l'equazione:

$$\begin{aligned} \tilde{\Lambda}x^{\max} = \mu^{\max}x^{\max} &\Rightarrow \sum_{m=1}^p \frac{w_l}{w_m} x_m^{\max} = px_l^{\max} \quad \text{per ogni } l \in P \Rightarrow \\ &\Rightarrow \frac{x_l^{\max}}{w_l} = \frac{1}{p} \sum_{m=1}^p \frac{x_m^{\max}}{w_m} \quad \text{per ogni } l \in P \end{aligned}$$

per cui x_l^{\max}/w_l non dipende da l : il vettore x^{\max} è proporzionale a w . ■

Osservazione 8 *Per trovare il vettore dei pesi w , non è necessario determinare l'autovettore dominante: basta normalizzare qualsiasi colonna di $\tilde{\Lambda}$.*

Dimostrazione.

$$\frac{\tilde{\lambda}_{lm}}{\sum_{l=1}^p \tilde{\lambda}_{lm}} = \frac{\frac{w_l}{w_m}}{\sum_{i=1}^p \frac{w_i}{w_m}} = \frac{w_l}{\sum_{i=1}^p w_i} = w_l$$

■

Il procedimento che ricava i pesi normalizzando gli elementi della colonna l corrisponde a usare un albero ricoprente costituito da una stella centrata nel vertice corrispondente all'indicatore f_l . In una matrice coerente, infatti, ogni albero ricoprente, e quindi ogni stella ricoprente, produce gli stessi pesi.

Esempio 51 A un decisore è stato chiesto di indicare i tassi di sostituzione fra tre funzioni di utilità normalizzata \tilde{u}_1 , \tilde{u}_2 e \tilde{u}_3 . Le sue risposte portano alla seguente matrice.

$$\tilde{\Lambda} = \begin{array}{c|ccc} & \tilde{u}_1 & \tilde{u}_2 & \tilde{u}_3 \\ \hline \tilde{u}_1 & 1 & 2 & 6 \\ \tilde{u}_2 & 1/2 & 1 & 3 \\ \tilde{u}_3 & 1/6 & 1/3 & 1 \end{array}$$

Si può verificare che la matrice è positiva, reciproca e coerente. Quindi, è possibile costruire a partire da essa una funzione di utilità additiva $u = \sum_{l=1}^p w_l \tilde{u}_l$. Il vettore dei pesi è $w = [0.6 \ 0.3 \ 0.1]^T$, e coincide, a meno di una costante moltiplicativa, con ciascuna delle colonne di $\tilde{\Lambda}$ (la costante è 6/10 per la prima colonna, 3/10 per la seconda e 1/10 per la terza).

Studiando il polinomio caratteristico di $\tilde{\Lambda}$

$$\begin{aligned} |\mu I - \tilde{\Lambda}| &= \begin{vmatrix} \mu - 1 & -2 & -6 \\ -1/2 & \mu - 1 & -3 \\ -1/6 & -1/3 & \mu - 1 \end{vmatrix} = (\mu - 1)^3 - 1 - 1 - (\mu - 1) - (\mu - 1) - (\mu - 1) = \\ &= \mu^3 - 3\mu^2 + 3\mu - 1 - 1 - 1 - 3\mu + 3 = \mu^3 - 3\mu^2 = \mu^2(\mu - 3) \end{aligned}$$

si verifica che due autovalori sono nulli e quello dominante è pari a $p = 3$. L'auto-vettore corrispondente risolve il sistema di equazioni:

$$(3I - \tilde{\Lambda})x = 0 \Rightarrow \begin{bmatrix} 3 - 1 & -2 & -6 \\ -1/2 & 3 - 1 & -3 \\ -1/6 & -1/3 & 3 - 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 2x_2 - 6x_3 \\ -1/2x_1 + 2x_2 - 3x_3 \\ -1/6x_1 - 1/3x_2 + 2x_3 \end{bmatrix} = 0$$

ed è proporzionale al vettore dei pesi w .

Esempio 52 A un decisore è stato chiesto di indicare i tassi marginali di sostituzione fra tre funzioni di utilità normalizzata \tilde{u}_1 , \tilde{u}_2 e \tilde{u}_3 . Le sue risposte portano alla seguente matrice.

$$\tilde{\Lambda} = \begin{array}{c|ccc} & \tilde{u}_1 & \tilde{u}_2 & \tilde{u}_3 \\ \hline u_1 & 1 & 2 & 4 \\ u_2 & 1/2 & 1 & 3 \\ u_3 & 1/4 & 1/3 & 1 \end{array}$$

La matrice è positiva e reciproca, ma non coerente, dato che $\tilde{\lambda}_{31} \neq \tilde{\lambda}_{32}\tilde{\lambda}_{21}$. Ogni colonna suggerisce un vettore di pesi diverso

$$w^{(1)} = \begin{bmatrix} 0.571 \\ 0.286 \\ 0.143 \end{bmatrix} \quad w^{(2)} = \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix} \quad w^{(3)} = \begin{bmatrix} 0.5 \\ 0.375 \\ 0.125 \end{bmatrix}$$

La differenza tra i tre vettori è rilevante, e potrebbe portare a scegliere soluzioni diverse.

In una matrice incoerente, ogni albero ricoprente produce la propria stima dei pesi, in genere diversa da quella basata sugli altri alberi.

7.1.1 La ricostruzione di matrici coerenti

Se gli attributi sono molti, nessun decisore umano riesce in pratica a mantenere una perfetta coerenza. I pesi costruiti da $p - 1$ confronti sono quindi inaffidabili. Si può invece costruire l'intera matrice dei confronti a coppie, valutare l'entità dell'incoerenza e costruire un vettore dei pesi più affidabile. Tale vettore w permette quindi di ricostruire una matrice di confronti coerente. Se si vuole impiegare il vettore dei pesi w per una decisione, occorre che la matrice dei confronti da esso generata sia la più vicina possibile a quella misurata. In altre parole, si vuole risolvere il seguente problema di minimizzazione:

$$\begin{aligned} \min_w & \|W - \tilde{\Lambda}\| \\ \sum_{l \in P} w_l &= 1 \\ w_l &\geq 0 \quad l \in P \end{aligned}$$

dove le incognite sono i pesi w_l , W indica la matrice costituita dai rapporti w_l/w_m , $\tilde{\Lambda}$ la matrice costituita dai rapporti stimati dal decisore e $\|W - \tilde{\Lambda}\|$ è la norma della matrice $W - \tilde{\Lambda}$.

Definizione 28 Con norma si intende qualsiasi funzione associ a una matrice un valore reale tale che:

- sia non negativo per ogni matrice:

$$\|M\| \geq 0 \text{ per ogni } M$$

- sia nullo se e solo se la matrice è nulla:

$$\|M\| = 0 \Leftrightarrow M = 0$$

- se tutti gli elementi della matrice vengono moltiplicati per un fattore costante, risulta moltiplicato per il valore assoluto di tale costante:

$$\|\alpha M\| = |\alpha| \|M\|$$

- soddisfi la diseguaglianza triangolare:

$$\|M_1 + M_2\| = \|M_1\| + \|M_2\|$$

Ci sono infinite diverse definizioni di norma che rispettano queste proprietà; ognuna fornisce una diversa soluzione al problema. Le più comuni in letteratura sono:

$$\begin{aligned} \|W - \tilde{\Lambda}\|_1 &= \sum_{l \in P} \sum_{m \in P} \left| \frac{w_l}{w_m} - \tilde{\lambda}_{lm} \right| \\ \|W - \tilde{\Lambda}\|_2 &= \sum_{l \in P} \sum_{m \in P} \left| \frac{w_l}{w_m} - \tilde{\lambda}_{lm} \right|^2 \\ \|W - \tilde{\Lambda}\|_\infty &= \max_{l \in P, m \in P} \left| \frac{w_l}{w_m} - \tilde{\lambda}_{lm} \right| \end{aligned}$$

In particolare, la norma $\|W - \tilde{\Lambda}\|_2$ è la norma euclidea, che dà luogo al classico *metodo dei minimi quadrati*.

L'idea di fondo di tutti questi metodi è partire dalle informazioni fornite dal decisore e modificarle il minimo necessario a raggiungere la coerenza. *Il valore ottimo della funzione obiettivo si può assumere come misura dell'incoerenza iniziale.* Infatti, è nullo se e solo se la matrice $\tilde{\Lambda}$ stimata è coerente.

Per tener conto dell'incertezza del decisore, è stato anche proposto di sostituire i valori stimati $\tilde{\lambda}_{lm}$ con intervalli e di definire la norma componendo le distanze fra ciascun rapporto w_l/w_m e l'intervallo corrispondente, anziché il valore stimato.

Il metodo dell'autovettore

Un metodo molto usato in letteratura, proposto da Saaty¹, il fondatore dell'Analisi Gerarchica cui è dedicata la Sezione 7.2, si basa sugli autovalori della matrice stimata $\tilde{\Lambda}$. Poiché essa è positiva, il teorema di Perron-Frobenius garantisce che l'autovalore dominante μ^{\max} (cioè quello di modulo massimo) sia reale e positivo. Se $\tilde{\Lambda}$ è anche reciproca (una condizione facile da imporre in pratica), $\mu^{\max} \geq p$ e inoltre $\mu^{\max} = p$ se e solo se la matrice è coerente.

L'incoerenza della matrice $\tilde{\Lambda}$ può essere quindi misurata come

$$\gamma = \frac{\mu^{\max} - p}{p - 1}$$

che è l'opposto del valor medio degli altri $p - 1$ autovalori, dato che la loro somma è pari alla traccia, cioè a p .

Se la matrice è debolmente incoerente, $\mu^{\max} \approx p$ e gli altri autovalori sono prossimi a zero. L'autovettore x^{\max} associato a μ^{\max} , una volta normalizzato, fornisce un vettore di pesi con il quale si può costruire una matrice coerente. Saaty ha anche suggerito di considerare accettabile un'incoerenza non superiore al 10% dell'incoerenza media ottenuta a partire da matrici casuali positive e reciproche della dimensione p desiderata. Tale incoerenza media è una funzione di p di cui sono stati tabulati i valori.

Questo metodo è stato severamente criticato in tempi più recenti perché si è dimostrato che può produrre matrici *dominate*, cioè che talvolta produce matrici i cui elementi che differiscono da quella stimata (tutti, uno per uno) più di una matrice coerente generata da altri metodi. Tale matrice alternativa è strettamente migliore di quella del metodo dell'autovettore per qualsiasi definizione di norma.

Esempio 53 Consideriamo la matrice di confronti a coppie:

$$\tilde{\Lambda} = \begin{array}{cc|ccc} & & u_1 & u_2 & u_3 \\ & u_1 & 1 & 2 & 4 \\ & u_2 & 1/2 & 1 & 3 \\ & u_3 & 1/4 & 1/3 & 1 \end{array}$$

i cui autovalori sono $\mu_1 = 3.108$, $\mu_2 = -0,0091 + 0,2348i$ e $\mu_3 = -0,0091 - 0,2348i$. L'incoerenza è data da $\gamma = 0.054$, piuttosto ridotta, e la matrice coerente suggerita dal metodo è:

$$\tilde{\Lambda} = \begin{array}{cc|ccc} & & \tilde{u}_1 & \tilde{u}_2 & \tilde{u}_3 \\ & \tilde{u}_1 & 1 & 1.747 & 4.579 \\ & \tilde{u}_2 & 0.572 & 1 & 2.621 \\ & \tilde{u}_3 & 0.218 & 0.381 & 1 \end{array}$$

Esempio 54 Consideriamo la matrice di confronti a coppie:

¹Thomas L. Saaty (1926-2017), matematico iraniano operante negli Stati Uniti.

$$\tilde{\Lambda} = \begin{array}{c|ccc} & \tilde{u}_1 & \tilde{u}_2 & \tilde{u}_3 \\ \hline \tilde{u}_1 & 1 & 2 & 1/2 \\ \tilde{u}_2 & 1/2 & 1 & 3 \\ \tilde{u}_3 & 2 & 1/3 & 1 \end{array}$$

i cui autovalori sono $\mu_1 = 3.7262$, $\mu_2 = -0.3631 + 1.6044i$ e $\mu_3 = -0.3631 - 1.6044i$.

L'incoerenza è data da $\gamma = 0.3631$, molto forte. Salta all'occhio il fatto che il primo attributo sia più rilevante del secondo e il secondo del terzo, ma il terzo sia più rilevante del primo, violando non solo la coerenza, ma addirittura la transitività.

La matrice coerente suggerita dal metodo dell'autovettore è

$$\tilde{\Lambda} = \begin{array}{c|ccc} & u_1 & u_2 & u_3 \\ \hline u_1 & 1 & 0.874 & 1.145 \\ u_2 & 1.145 & 1 & 1.310 \\ u_3 & 0.874 & 0.763 & 1 \end{array}$$

ma in questo caso si può dubitare che si tratti di una matrice significativa.

7.1.2 Coerenza ad ogni costo?

La sostituzione della matrice dei confronti a coppie data $\tilde{\Lambda}$ in una matrice coerente W è generalmente un'operazione che distorce le informazioni fornite dal decisore. Parte della letteratura riporta che la nuova matrice W può risultare più gradita al decisore di quella da lui fornita, come se l'operazione lo aiutasse a raggiungere una coerenza di cui non era capace, ma che corrisponde alle sue reali aspirazioni. Parte della letteratura, invece, riporta che la matrice W può risultare meno gradita, come se l'operazione contraddicesse le preferenze del decisore in qualche aspetto importante.

Al momento, l'unica conclusione assodata è che conviene compiere questa operazione in modo interattivo anziché a scatola nera, magari procedendo per iterazioni successive. Cioè sottponendo ogni volta i risultati al decisore per conferma, e chiedendogli di modificare le sue indicazioni per avvicinarsi alla coerenza.

7.2 Analisi Gerarchica

L'*Analisi Gerarchica* o *Analytic Hierarchy Process* (*AHP*) nasce nel 1980, sulla spinta delle seguenti critiche avanzate da Saaty ai metodi classici per la costruzione della funzione di utilità:

1. la ricostruzione delle singole funzioni di utilità normalizzata è soggetta a forti errori di approssimazione;
2. la stima dei pesi è soggetta a forti errori di approssimazione quando il numero degli attributi p è elevato;
3. i vari errori di approssimazione si compongono a cascata.

Le approssimazioni nelle funzioni di utilità normalizzate sono specialmente forti nel caso di attributi qualitativi, per i quali è facile fornire un ordinamento fra impatti, ma molto più difficile misurare la forza relativa della preferenza fra i diversi impatti. Se le funzioni di utilità normalizzata non sono corrette, però, le successive operazioni di stima dei pesi e combinazione lineare per valutare l'utilità complessiva perdono significato. Infatti, qualsiasi alternativa paretiana di un problema può risultare

vincente se si sceglie un'opportuna funzione di utilità. Queste difficoltà pratiche rischiano di ridurre i metodi classici a “number-crunching”, cioè manipolazione di numeri senza senso concreto e senza controllo da parte del decisore.

L'*AHP* mira a definire in maniera più semplice e naturale la misura di preferenza fra impatti alternativi. I suoi elementi caratteristici sono:

1. la valutazione dell'utilità delle soluzioni rispetto a ciascun attributo tramite confronti a coppie anziché misure assolute dirette;
2. l'uso di scale qualitative anziché quantitative;
3. la valutazione dei pesi degli attributi tramite confronti a coppie;
4. la strutturazione degli attributi in una gerarchia;
5. la ricombinazione moltiplicativa dei pesi ai vari livelli della gerarchia.

Anche se non ha propriamente senso parlare di indipendenza preferenziale e di regolarità degli incrementi, le operazioni impiegate dall'*AHP* assumono implicitamente che i vari indicatori contribuiscano in modo additivo all'utilità complessiva di ogni soluzione. In effetti, i concetti fondamentali dell'*AHP* si possono interpretare come una versione approssimata e qualitativa, ma più robusta agli errori di modellazione, dei concetti fondamentali della teoria classica.

7.2.1 Determinazione delle utilità tramite confronti a coppie

Fissato un indicatore f_l , i metodi classici cercano di ricostruire l'intero andamento assoluto dell'utilità come funzione $\tilde{u}_l(f_l)$. La critica di Saaty si basa su considerazioni di psicologia cognitiva secondo le quali è molto difficile per un decisore umano assegnare ai valori degli indicatori dei valori quantitativi di utilità. L'*AHP* si concentra quindi su coppie di soluzioni e sulla forza della preferenza relativa fra i valori di un indicatore nelle due soluzioni della coppia. Si costruisce una matrice $\Lambda_l = \{\lambda_{xy}^{(l)}\}$ in cui ciascun elemento $\lambda_{xy}^{(l)}$ è associato a una coppia di alternative $(x, y) \in X \times X$ e valuta di quanto $f_l(x)$ sia preferibile a $f_l(y)$. Quindi, non si tratta di ricostruire l'intera funzione di utilità normalizzata rispetto a un indicatore su tutto l'insieme di valori che esso può assumere, ma di valutare dei rapporti di forza fra coppie di valori alternativi dati, cioè qualcosa di simile a dei rapporti di utilità $\tilde{u}_l(f_l(x)) / \tilde{u}_l(f_l(y))$.

Siccome richiede l'enumerazione esplicita di tutte le coppie di soluzioni, l'*Analisi Gerarchica* si può applicare solo a problemi finiti e con un numero piccolo di soluzioni, oppure a problemi in cui l'insieme delle soluzioni è stato preliminarmente sfrondato sino a ridurlo a un piccolo insieme finito.

7.2.2 Scale qualitative

La seconda idea fondamentale dell'Analisi Gerarchica è di semplificare la valutazione misurando la preferenza fra i due valori $f_l(x)$ e $f_l(y)$ non con una scala quantitativa, ma con una scala qualitativa, detta *scala di Saaty*:

1. $\lambda_{xy}^{(l)} = 1$: $f_l(x)$ e $f_l(y)$ sono *indifferenti*;
2. $\lambda_{xy}^{(l)} = 3$: $f_l(x)$ è *poco* preferibile a $f_l(y)$;
3. $\lambda_{xy}^{(l)} = 5$: $f_l(x)$ è *abbastanza* preferibile a $f_l(y)$;

4. $\lambda_{xy}^{(l)} = 7$: $f_l(x)$ è molto preferibile a $f_l(y)$;
5. $\lambda_{xy}^{(l)} = 9$: $f_l(x)$ è assolutamente preferibile a $f_l(y)$.

I pesi 2, 4, 6 e 8 sono utilizzati per valutazioni intermedie. Il simbolo λ viene usato per analogia con il tasso marginale di sostituzione, ma si tratta chiaramente di una misura molto più empirica e approssimata. I valori della scala di Saaty sono assolutamente arbitrari, ma derivano da studi psicologici, secondo i quali è inappropriato chiedere a un decisore di indicare valori quantitativi esatti per i rapporti di preferenza, dato che gli esseri umani non sono in grado di discriminare più di 5 – 9 livelli in modo sensato e coerente.

Inoltre, l'uso di una scala qualitativa consente di confrontare quantità disomogenee e non espresse in modo quantitativo, traducendo dei giudizi verbali in valori numerici. L'idea è di dare per scontato che i valori numerici utilizzati siano solo una rozza approssimazione della realtà; saperlo mette automaticamente in guardia dal fidarsene troppo, rischio che invece si corre facilmente con le utilità normalizzate stimate dai metodi classici. Inoltre, costruendo un'intera matrice di confronti a coppie, è possibile valutarne subito l'incoerenza e rimodularla a posteriori fino a renderla coerente, invece di forzare fin da subito una coerenza fittizia, che potrebbe introdurre errori non emendabili nelle fasi successive.

I metodi descritti nella Sezione 7.1.1 consentono di derivare dalla matrice dei confronti a coppie una matrice coerente, e da questa un vettore u_l che associa ai singoli valori considerati per l'indicatore, cioè ad ogni soluzione $x \in X$, un valore numerico u_{xl} positivo e normalizzato.

Definizione 29 Definiamo matrice di valutazione la matrice $U = \{u_{xl}\}$ che contiene la valutazione u_{xl} di ciascuna alternativa $x \in X$ rispetto a ciascun indicatore $l \in P$, ottenuta partendo dalla matrice dei confronti a coppie Λ_l .

Le valutazioni u_{xl} così ottenute sostituiscono i valori delle funzioni di utilità normalizzata $\tilde{u}_l(f_l(x))$ della teoria classica. Anche questi valori ricadono in $[0, 1]$ e sono maggiori per le alternative preferite, ma in generale i valori pessimo e ottimo di un indicatore non corrisponderanno più ai valori estremi 0 e 1.

7.2.3 Determinazione dei pesi con confronti a coppie

Anche i pesi degli attributi vengono costruiti partendo da matrici di confronti a coppie riempite di valori qualitativi tratti dalla scala di Saaty. Il confronto fra due attributi punta a valutare il peso relativo dell'uno rispetto all'altro. Con i metodi descritti nella Sezione 7.1.1 si può ricavare dalla matrice dei confronti a coppie fra attributi, un vettore di pesi normalizzati w_l associati agli attributi del problema. Questo vettore non risulta quindi dall'identificazione di coppie di impatti indifferenti, ma dall'elaborazione dei giudizi sul peso relativo degli attributi forniti dal decisore. Il significato pratico di questi pesi, però, è lo stesso: essi vengono usati per combinare in modo additivo le valutazioni u_{xl} , in modo da ottenere la valutazione complessiva $u_x = \sum_{l=1}^P u_{xl}$ di ciascuna alternativa $x \in X$, in base alla quale la scelta verrà compiuta.

7.2.4 Strutturazione gerarchica degli attributi

Costruire una matrice di confronti a coppie completa fra tutti gli attributi di un problema può essere impraticabile se gli attributi sono molti. Come si è anticipato nel Capitolo 56, però, si possono raccogliere gli attributi del problema in categorie omogenee, strutturate in modo gerarchico:

1. al livello delle foglie compaiono gli attributi elementari;
2. ai livelli superiori compaiono attributi riassuntivi, via via più generali;
3. alla radice compare una specie di obiettivo globale.

Ogni attributo del problema, cioè, viene riassunto in un obiettivo più generale e scomposto in obiettivi più particolari. Ad esempio, nella Valutazione di Impatto Ambientale (VIA) l'obiettivo generale di soddisfazione del decisore viene spesso scomposto nei macro-obiettivi di garantire un impatto positivo nei settori ambientale, economico e sociale. A sua volta, l'impatto ambientale può essere scomposto nei settori acque, terreno, aria, rumore; il settore "aria" nei diversi generi di inquinanti, questi nelle diverse aree geografiche, e così via. La Figura 7.1 illustra un esempio (semplicificato) relativo alla metrotranvia nel comune di Como, dove ambiente, economia e società erano accompagnati dal settore aggiuntivo del traffico.

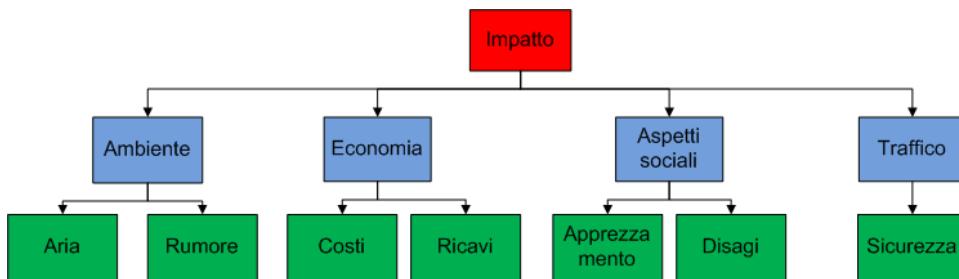


Figura 7.1: La struttura gerarchica del problema di VIA per la metrotranvia di Como (versione fortemente semplificata)

I vantaggi della strutturazione gerarchica degli attributi sono:

1. si confrontano fra loro solo attributi omogenei (ad esempio, l'inquinamento da CO con quello da NO_x , e non con l'aumento di manodopera o i costi finanziari);
2. si può affidare ciascun insieme di confronti a coppie a un *sottodecisore esperto* di un determinato campo, in modo da ottenere indicazioni più sensate;
3. il numero totale dei confronti a coppie diminuisce fortemente.

Lo svantaggio fondamentale è che i confronti a coppie al livello delle foglie considerano indicatori concreti e ben definiti, mentre quelli ai livelli superiori considerano obiettivi astratti e generali, per i quali non ha nemmeno senso una misura quantitativa. Per esempio, al livello delle foglie si confronta il peso dell'inquinamento da CO con quello dell'inquinamento da NO_x , mentre ai livelli superiori si confronta l'impatto ambientale di un'opera con l'impatto sociale. Questo secondo confronto, per quanto sia espresso con una scala numerica, è ovviamente qualitativo. Tipicamente, i livelli superiori sono assegnati ai decisori politici, in quanto esperti di compromessi fra temi generali.

7.2.5 Ricomposizione gerarchica

L'uso di una struttura gerarchica complica leggermente la determinazione dei pesi degli indicatori. Infatti, solo gli indicatori figli dello stesso nodo nell'albero vengono confrontati fra loro a coppie. Questo dà luogo a pesi che sono normalizzati all'interno

di ciascun gruppo di figli, ma non sono confrontabili con quelli degli altri gruppi. La struttura ad albero, però, consente di costruire il vettore dei pesi degli attributi in modo progressivo risalendo di livello in livello dalle foglie sino alla radice dell'albero. Ad ogni livello, si costruisce una matrice di confronti a coppie fra tutti i nodi figli dello stesso padre, e se ne deriva un vettore di pesi coi metodi già descritti. I pesi in ogni vettore hanno somma unitaria, e descrivono il peso relativo degli uni rispetto agli altri. Per confrontare tali pesi con quelli degli altri nodi dell'intero albero è sufficiente rinormalizzarli in modo che la loro somma coincida col peso del nodo padre, cioè semplicemente moltiplicarli per il peso del nodo padre. Qualora l'albero abbia parecchi livelli, è sufficiente moltiplicare il peso di ciascun attributo foglia per quello di tutti i nodi incontrati lungo il percorso che conduce alla radice. In questo modo, la somma di tutti i pesi dei nodi su un livello è sempre unitaria; in particolare, la somma dei pesi w_l dei nodi foglia è pari a 1.

Ricostruiti i pesi w_l , la matrice di valutazione u_{xl} fornisce per ogni alternativa e ogni indicatore elementare una valutazione. Come nella teoria classica, si ricava l'utilità complessiva $u(x)$ di ciascuna alternativa x eseguendo la combinazione convessa delle utilità con i pesi. Questo corrisponde a moltiplicare la matrice di valutazione U per il vettore dei pesi w :

$$u_x = \sum_{l=1}^p u_{xl} w_l \Leftrightarrow u = U \cdot w$$

I valori delle utilità così ottenuti vengono usati per ordinare le alternative e risolvere il problema di decisione.

Esempio 55 Consideriamo l'acquisto di una casa. Supponiamo che vi siano tre alternative, denominate a_1 , a_2 e a_3 e che la scelta dipenda da quattro attributi (prezzo, dimensione, qualità della vita nella zona, condizioni di pagamento). La gerarchia del problema vede quindi una radice, che corrisponde all'obiettivo globale, da cui derivano quattro nodi foglia (gli attributi).

Relativamente a ciascuno dei quattro attributi, la fase di confronti a coppie fra le alternative porta alle seguenti matrici, e quindi ai seguenti vettori di pesi (essendoci poche alternative, si suppone di aver ottenuto matrici coerenti, dunque indipendenti dal metodo usato per costruire le utilità):

1. prezzo

$$\Lambda_p = \begin{bmatrix} 1 & 1/9 & 1/8 \\ 9 & 1 & 9/8 \\ 8 & 8/9 & 1 \end{bmatrix} \Rightarrow u_p = \begin{bmatrix} 1/18 \\ 9/18 \\ 8/18 \end{bmatrix}$$

2. dimensioni

$$\Lambda_d = \begin{bmatrix} 1 & 9 & 9 \\ 1/9 & 1 & 1 \\ 1/9 & 1 & 1 \end{bmatrix} \Rightarrow u_d = \begin{bmatrix} 9/11 \\ 1/11 \\ 1/11 \end{bmatrix}$$

3. qualità della vita nella zona

$$\Lambda_q = \begin{bmatrix} 1 & 1/9 & 1/4 \\ 9 & 1 & 9/4 \\ 4 & 4/9 & 1 \end{bmatrix} \Rightarrow u_q = \begin{bmatrix} 1/14 \\ 9/14 \\ 4/14 \end{bmatrix}$$

4. condizioni di pagamento

$$\Lambda_c = \begin{bmatrix} 1 & 3 & 3/5 \\ 1/3 & 1 & 1/5 \\ 5/3 & 5 & 1 \end{bmatrix} \Rightarrow u_c = \begin{bmatrix} 3/9 \\ 1/9 \\ 5/9 \end{bmatrix}$$

Supponiamo che secondo il decisore i quattro attributi abbiano lo stesso peso.

$$\Lambda_w = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \Rightarrow w = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

Avendo l'albero un solo livello, non occorre fare alcuna ricombinazione dei pesi. Basta costruire la matrice di valutazione, accostando i vettori che esprimono l'utilità delle alternative rispetto agli attributi

$$U = [u_p \mid u_d \mid u_q \mid u_c]$$

ed eseguire la combinazione convessa delle utilità con i pesi, moltiplicando la matrice di valutazione per il vettore dei pesi.

$$u = Uw = \begin{bmatrix} 1/18 & 9/11 & 1/14 & 3/9 \\ 9/18 & 1/11 & 9/14 & 1/9 \\ 8/18 & 1/11 & 4/14 & 5/9 \end{bmatrix} \cdot \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.320 \\ 0.336 \\ 0.344 \end{bmatrix}$$

che porta a scegliere l'alternativa a_3 , anche se le altre due sono poco distanti. Se i coefficienti assegnati durante i confronti a coppie non erano molto affidabili, questo suggerisce che sia opportuno fare un'analisi di sensitività per valutare eventuali variazioni, come nell'Esempio 31 della Sezione 5.6.3.

Se l'albero ha più livelli, è del tutto equivalente valutare le alternative in ciascun nodo dell'albero, cioè rispetto al corrispondente indicatore di livello superiore: basta risalire l'albero interrompendo la ricombinazione nel nodo che corrisponde all'attributo considerato. Per esempio, se si vogliono conoscere le utilità ambientali delle alternative, basta risalire l'albero sino al nodo Ambiente, e rinormalizzare i pesi dei soli attributi ambientali moltiplicandoli solo per i pesi dei nodi lungo il cammino dalle foglie a tale nodo e poi eseguire la combinazione convessa delle sole utilità per indicatori ambientali con tali pesi.

7.2.6 Ibridazioni con la teoria dell'utilità

La strutturazione gerarchica degli attributi è un'innovazione dell'AHP che è stata poi riadattata al *corpus* dei metodi classici. Può infatti avvenire che si abbia fiducia nelle funzioni di utilità normalizzata stimate, ma un numero di indicatori troppo alto per determinarne i pesi confrontandoli tutti fra loro in modo sensato. In questo caso, è possibile accettare i valori di utilità normalizzata, ma stimare i pesi con i confronti a coppie, le scale qualitative e la ricombinazione moltiplicativa proposte dall'Analisi Gerarchica.

Esempio 56 Consideriamo un esempio molto semplificato relativo al caso di studio della metrotranvia di Como introdotto nel Capitolo . Per semplicità consideriamo un numero limitato di alternative e di attributi, ma manteniamo un albero degli attributi a più livelli. Le alternative scelte sono:

0. Alternativa zero: nessun intervento;
1. Navetta Ferroviaria;
2. Interoperabile (avendo scelto uno dei percorsi);

Sottomatrice	Settore	Sottosettore	Alternative			
			0	1	2	3
U_a^T	Ambiente	Aria	0.0	0.1	0.6	0.8
		Rumore	0.4	0.2	0.4	0.6
		Vibrazioni	0.4	0.2	0.4	0.6
		Paesaggio	1.0	1.0	0.8	0.5
		Assetto territ.	0.0	0.3	1.0	0.6
U_e^T	Economia	Costi	1.0	0.9	0.6	0.0
		Ricavi	0.0	0.3	0.7	1.0
		Valore immobili	0.0	0.5	1.0	1.0
U_s^T	Società	Apprezzamento	0.0	0.2	0.7	0.9
		Disagi	1.0	1.0	0.7	0.2
		Accessibilità	0.0	0.1	0.7	0.9
		Occupazione	0.0	0.5	1.0	1.0
		Effetti indotti	0.0	0.3	0.6	1.0
U_t^T	Trasporti	Sicurezza	0.0	0.2	0.6	1.0
		Congestione	0.0	0.4	0.9	1.0
		Interferenze	1.0	0.8	0.6	0.2

Tabella 7.1: Matrice di valutazione (trasposta) per il problema (semplificato) della metrotranvia di Como

3. Tramvia Anulare.

Eseguita la stima degli impatti per ogni alternativa e la costruzione delle funzioni di utilità normalizzata, si suppone di ottenere la matrice di valutazione U riportata in Tabella 7.1. La matrice è trasposta per comodità di rappresentazione e divisa gerarchicamente in sottomatrici. Ogni colonna della matrice (riga nella tabella) fornisce l'utilità normalizzata per un indicatore, con valori compresi fra 0 (per la prestazione pessima) e 1 (per quella ottima).

Si può quindi procedere costruendo i pesi degli attributi con confronti a coppie e valutazioni qualitative. In una prima fase, quattro decisori tecnici valutano l'importanza relativa dei sottosettori raccolti in ciascun settore. Supponiamo che le matrici dei confronti a coppie per ciascun settore siano quelle rappresentate nelle Tabelle 7.2, 7.3, 7.4 e 7.5.

Λ_a	Aria	Rumore	Vibrazioni	Paesaggio	Assetto Territ.
Aria	1	1	1	5/4	1
Rumore	1	1	1	3/4	1
Vibrazioni	1	1	1	3/4	3/4
Paesaggio	4/5	4/3	4/3	1	1
Assetto Territ.	1	1	4/3	1	1

Tabella 7.2: Matrice dei confronti a coppie per i sottosettori del settore Ambiente

Da ciascuna matrice si deriva il corrispondente vettore dei pesi dei sottosettori. Siccome i decisori non sono del tutto coerenti, ogni matrice va prima resa coerente, per poi ricavarne il vettore dei pesi. Questo è stato fatto con il metodo dei minimi quadrati. I quattro vettori risultanti sono riportati nella Tabella 7.6 e aggiunti alla struttura gerarchica complessiva nella Tabella 7.7.

In una seconda fase, un decisore politico valuta l'importanza relativa fra i settori, di nuovo tramite confronti a coppie. La Tabella 7.8 riporta i confronti stessi.

Λ_e	Costi	Ricavi	Valore Immob.
Costi	1	5/4	5/4
Ricavi	4/5	1	1
Valore Immob.	4/5	1	1

Tabella 7.3: Matrice dei confronti a coppie per i sottosettori del settore Economia

Λ_s	Apprezz.	Disagi	Access.	Occup.	Eff. Ind.
Apprezzamento	1	1	1	5/4	1
Disagi	1	1	5/4	3/4	1
Accessibilità	1	4/5	1	3/2	1
Occupazione	4/5	4/3	2/3	1	3/4
Effetti Indotti	1	1	1	4/3	1

Tabella 7.4: Matrice dei confronti a coppie per i sottosettori del settore Società

Λ_t	Sicurezza	Congestione	Interferenze
Sicurezza	1	3/4	1
Congestione	4/3	1	5/4
Interferenze	1	4/5	1

Tabella 7.5: Matrice dei confronti a coppie per i sottosettori del settore Trasporti

Ambiente		Economia		Società		Trasporti	
Sottosett.	w_a	Sottosett.	w_e	Sottosett.	w_s	Sottosett.	w_t
Aria	0.209	Costi	0.384	Apprezz.	0.209	Sicur.	0.301
Rumore	0.188	Ricavi	0.308	Disagi	0.197	Congest.	0.392
Vibrazioni	0.178	Val. Imm.	0.308	Access.	0.207	Interf.	0.307
Paesaggio	0.214			Occup.	0.176		
Ass. Terr.	0.211			Eff. Ind.	0.211		

Tabella 7.6: Vettori dei pesi per i sottosettori dei quattro settori

Sottomatrice	Settore	Sottosettore	Alternative			
			0	1	2	3
U_a^T	Ambiente	Aria	0.209	0.0	0.1	0.6
		Rumore	0.188	0.4	0.2	0.4
		Vibrazioni	0.178	0.4	0.2	0.4
		Paesaggio	0.214	1.0	1.0	0.8
		Assetto Territ.	0.211	0.0	0.3	1.0
U_e^T	Economia	Costi	0.384	1.0	0.9	0.6
		Ricavi	0.308	0.0	0.3	0.7
		Val. Imm.	0.308	0.0	0.5	1.0
U_s^T	Società	Apprezz.	0.209	0.0	0.2	0.7
		Disagi	0.197	1.0	1.0	0.7
		Access.	0.207	0.0	0.1	0.7
		Occup.	0.176	0.0	0.5	1.0
		Eff. Ind.	0.211	0.0	0.3	0.6
U_t^T	Trasporti	Sicurezza	0.301	0.0	0.2	0.6
		Congestione	0.392	0.0	0.4	0.9
		Interferenze	0.307	1.0	0.8	0.6

Tabella 7.7: Matrice di valutazione (trasposta) per il problema (semplificato) della metrotranvia di Como, con i pesi calcolati per i sottosettori dei quattro settori

	Ambiente	Economia	Società	Trasporti
Ambiente	1	3/2	3/4	1
Economia	2/3	1	1/2	3/4
Società	4/3	2	1	1
Trasporti	1	4/3	1	1

Tabella 7.8: Matrice dei confronti a coppie fra i quattro settori

Se ne ricavano i pesi $w^{(1)}$ dei settori al livello 1 dell'albero (vedi Tabella 7.9), ancora una volta applicando il metodo dei minimi quadrati per risolvere le incoerenze del decisore. La Tabella 7.10 riporta la struttura ad albero complessiva con i pesi a entrambi i livelli.

Settori	$w^{(1)}$
Ambiente	0.252
Economia	0.173
Società	0.312
Trasporti	0.263

Tabella 7.9: Vettore dei pesi dei quattro settori

Sottomatrice	Settore	Sottosettore	Alternatives			
			0	1	2	3
U_a^T	Ambiente 0.252	Aria	0.209	0.0	0.1	0.6
		Rumore	0.188	0.4	0.2	0.4
		Vibrazioni	0.178	0.4	0.2	0.4
		Paesaggio	0.214	1.0	1.0	0.8
		Assetto Territ.	0.211	0.0	0.3	1.0
U_e^T	Economia 0.173	Costi	0.384	1.0	0.9	0.6
		Ricavi	0.308	0.0	0.3	0.7
		Val. Imm.	0.308	0.0	0.5	1.0
U_s^T	Società 0.312	Apprezz.	0.209	0.0	0.2	0.7
		Disagi	0.197	1.0	1.0	0.7
		Access.	0.207	0.0	0.1	0.7
		Occup.	0.176	0.0	0.5	1.0
		Eff. Ind.	0.211	0.0	0.3	0.6
U_t^T	Trasporti 0.263	Sicurezza	0.301	0.0	0.2	0.6
		Congestione	0.392	0.0	0.4	0.9
		Interferenze	0.307	1.0	0.8	0.6

Tabella 7.10: Matrice di valutazione (trasposta) per il problema (semplificato) della metrotranvia di Como, con i pesi calcolati per i sottosettori e per i quattro settori

A questo punto, sono possibili due vie equivalenti. La prima consiste nel valutare i pesi di ciascun sottosettore moltiplicando il suo peso relativo al settore di appartenenza (Tabella 7.6) per il peso del settore, e via via per ogni altro peso incontrato lungo il cammino verso la radice dell'albero degli attributi (in questo caso, i livelli sono solo due). Questo fornisce un vettore complessivo di $5 + 3 + 5 + 3 = 16$ pesi di somma unitaria, rappresentato nella Tabella 7.11.

Con questi pesi si possono combinare le utilità della Tabella 7.1, moltiplicando la matrice di valutazione complessiva per il vettore complessivo. Questo produce le utilità della Tabella 7.13.

La seconda via consente di valutare anche le utilità parziali di ciascuna alternativa rispetto ai singoli settori. Per farlo, si risale gradualmente l'albero, moltiplicando le sottomatrici di utilità per i sottovettori dei pesi normalizzati a ciascun livello. In questo modo, si ottengono le utilità delle alternative normalizzate rispetto a ciascun settore. La Tabella 7.12 riporta la matrice $U^{(1)}$ delle utilità aggregate al primo livello dell'albero.

Sectors	w	Alternative			
		0	1	2	3
Aria	0.053	0.0	0.1	0.6	0.8
Rumore	0.047	0.4	0.2	0.4	0.6
Vibrazioni	0.045	0.4	0.2	0.4	0.6
Paesaggio	0.054	1.0	1.0	0.8	0.5
Assetto terr.	0.053	0.0	0.3	1.0	0.6
Costi	0.066	1.0	0.9	0.6	0.0
Ricavi	0.053	0.0	0.3	0.7	1.0
Valore immobili	0.053	0.0	0.5	1.0	1.0
Apprezzamento	0.065	0.0	0.2	0.7	0.9
Disagi	0.061	1.0	1.0	0.7	0.2
Accessibilità	0.065	0.0	0.1	0.7	0.9
Occupazione	0.055	0.0	0.5	1.0	1.0
Effetti indotti	0.066	0.0	0.3	0.6	1.0
Sicurezza	0.079	0.0	0.2	0.6	1.0
Congestione	0.103	0.0	0.4	0.9	1.0
Interferenze	0.081	1.0	0.8	0.6	0.2

Tabella 7.11: Matrice di valutazione (trasposta) per il problema (semplificato) della metrotranvia di Como con il vettore dei pesi complessivo per tutti gli attributi

Submatrix	Sector	Alternative			
		0	1	2	3
U_a^T	Ambiente 0.252	0.360	0.371	0.654	0.620
U_e^T	Economia 0.173	0.385	0.593	0.755	0.616
U_s^T	Società 0.312	0.197	0.411	0.732	0.801
U_t^T	Trasporti 0.263	0.307	0.463	0.718	0.754

Tabella 7.12: Matrice di valutazione delle alternative aggregata rispetto ai quattro settori

Da qui si può ottenere l'utilità globale (cioè alla radice dell'albero) di ciascuna alternativa, moltiplicando la matrice aggregata al primo livello per il vettore dei pesi dei settori, ottenendo ancora una volta i valori riportati nella Tabella 7.13.

Utilità $u = U^{(1)}w^{(1)}$	Alternative			
	0	1	2	3
	0.299	0.446	0.713	0.711

Tabella 7.13: Vettore delle utilità complessivo

7.2.7 Rank reversal

Il maggior difetto dell'Analisi Gerarchica, che ha dato luogo a un vivace dibattito di natura anche filosofica, sta nel fenomeno dell'*inversione delle preferenze (rank reversal)*: l'ordinamento fra le alternative dipende sostanzialmente da quali alternative sono presenti. Questo perché le valutazioni u_{xl} non sono utilità assolute, ma derivano da confronti a coppie, e quindi dipendono dai valori effettivamente confrontati.

Lo stesso problema si presenta nel campionato di calcio, che del resto è un sistema di ordinamento basato su confronti a coppie, anche se il calcolo della funzione di ordinamento è condotto diversamente. Supponiamo che le squadre A e B risultino vincenti su insiemi diversi di altre squadre e che A sia leggermente superiore a B ; se compare una nuova squadra C , dominata da entrambe, ma che si comporta meglio negli scontri diretti con A che in quelli con B , questo provoca una redistribuzione dei pesi che può portare alla vittoria finale di B . Fenomeni analoghi si presentano negli scontri elettorali, dove la presenza di un candidato destinato alla sconfitta può però danneggiare in misura diversa i candidati più quotati, modificando il risultato finale dell'elezione.

Il problema si presenta anche nel caso di matrici di confronto a coppie perfettamente coerenti. In particolare, sono stati proposti esempi in cui persino lo sdoppiamento di un'alternativa in due alternative perfettamente identiche, che produce un problema del tutto equivalente a quello di partenza, modifica l'ordinamento finale. Ovviamente, in una situazione del genere, le due alternative identiche dovrebbero semplicemente presentarsi appaiate senza alterare l'ordinamento delle altre.

Esempio 57 Riprendiamo l'Esempio 55, che riguardava l'acquisto di una casa. Senza alcuna variazione nei confronti fra le tre alternative, introduciamo una quarta alternativa a_4 . Le nuove matrici di confronto a coppie sono identiche alle precedenti, salvo che sono orlate con una riga e una colonna addizionali, che riportano i confronti che riguardano a_4 .

1. prezzo

$$\Lambda_p = \begin{bmatrix} 1 & 1/9 & 1/8 & 1/4 \\ 9 & 1 & 9/8 & 9/4 \\ 8 & 8/9 & 1 & 2 \\ 4 & 4/9 & 1/2 & 1 \end{bmatrix} \Rightarrow u_p = \begin{bmatrix} 1/22 \\ 9/22 \\ 8/22 \\ 4/22 \end{bmatrix}$$

2. dimensioni

$$\Lambda_d = \begin{bmatrix} 1 & 9 & 9 & 9 \\ 1/9 & 1 & 1 & 1 \\ 1/9 & 1 & 1 & 1 \\ 1/9 & 1 & 1 & 1 \end{bmatrix} \Rightarrow u_d = \begin{bmatrix} 9/12 \\ 1/12 \\ 1/12 \\ 1/12 \end{bmatrix}$$

3. qualità della vita nella zona

$$\Lambda_q = \begin{bmatrix} 1 & 1/9 & 1/4 & 1/8 \\ 9 & 1 & 9/4 & 9/8 \\ 4 & 4/9 & 1 & 1/2 \\ 8 & 8/9 & 2 & 1 \end{bmatrix} \Rightarrow u_q = \begin{bmatrix} 1/22 \\ 9/22 \\ 4/22 \\ 8/22 \end{bmatrix}$$

4. condizioni di pagamento

$$\Lambda_c = \begin{bmatrix} 1 & 1/3 & 3/5 & 3/5 \\ 1/3 & 1 & 1/5 & 1/5 \\ 5/3 & 5 & 1 & 1 \\ 5/3 & 5 & 1 & 1 \end{bmatrix} \Rightarrow u_c = \begin{bmatrix} 3/14 \\ 1/14 \\ 5/14 \\ 5/14 \end{bmatrix}$$

La nuova alternativa sottrae alle precedenti parte dell'utilità per ciascuno degli attributi. Le sue prestazioni sono piuttosto scarse sui primi due attributi, mentre si classifica seconda per l'attributo "qualità" e prima a pari merito con a_3 per l'attributo "condizioni". La nuova matrice di valutazione e l'utilità globale di ciascun'alternativa sono le seguenti:

$$u = Uw = \begin{bmatrix} 1/22 & 9/12 & 1/22 & 3/14 \\ 9/22 & 1/12 & 9/22 & 1/14 \\ 8/22 & 1/12 & 4/22 & 5/14 \\ 4/22 & 1/12 & 8/22 & 5/14 \end{bmatrix} \cdot \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.264 \\ 0.243 \\ 0.246 \\ 0.246 \end{bmatrix}$$

che porta a scegliere l'alternativa a_1 . L'alternativa a_4 , infatti, comportandosi male rispetto al secondo attributo (dove a_1 primeggia) e bene rispetto al terzo (dove a_2 primeggia) e al quarto (dove a_3 primeggia), in pratica apre la strada ad a_1 .

Il problema del *rank reversal* è ineliminabile se non si assume una scala assoluta di valori.

7.2.8 Scale absolute o metodo della stima a priori

Per ovviare al *rank reversal*, Saaty propose un sistema basato sul confronto non fra le alternative singole, ma fra classi predefinite di alternative.

Ad esempio, dovendo giudicare dei candidati a un posto di lavoro per il quale si richiede la conoscenza dell'inglese e dell'uso del computer, non si confronteranno a coppie i candidati riguardo la loro competenza su questi due aspetti, ma si stabiliranno dei livelli di competenza (ad esempio, "scarsa", "sufficiente", "discreta", "buona" e "ottima") e si confronteranno a coppie questi livelli. Così una preparazione ottima in inglese potrà valere 7 volte una preparazione discreta, mentre una preparazione ottima nell'uso del computer potrà valere 3 volte una discreta, e così via. Quindi, si tratterà di assegnare ogni alternativa (cioè ogni candidato) a una delle classi.

Il metodo ha l'ulteriore vantaggio di ridurre il numero di confronti richiesti, dato che la matrice ha una dimensione determinata dal numero di classi, e non di alternative. Questo consente di applicare l'*AHP* anche a problemi con molte alternative. Siccome il metodo consente di valutare preliminarymente le utilità di ciascuna classe, si può applicare persino quando le alternative non sono tutte note sin dal principio. Per esempio, se i candidati si presentano in modo distribuito nel tempo, ma vanno valutati immediatamente, senza conoscere i candidati successivi. In tale situazione, stabilire che il candidato A è 5 volte meglio del candidato B è impossibile se il candidato B non è ancora noto, e potrebbe essere difficile anche

nel caso di un candidato già valutato, perché un eventuale candidato successivo potrebbe rendere tale rapporto troppo piccolo o troppo grande.

Lo svantaggio del metodo sta nell'arbitrarietà delle classi: spostando il confine di una classe anche di poco, alcune alternative cambiano classe e la loro valutazione può cambiare di molto.

Esempio 58 Supponiamo di dover assumere un collaboratore. Vi sono quattro criteri di scelta: studi, curriculum, esperienza e conoscenza dell'inglese. A priori, non sappiamo quanti candidati si presenteranno. Di conseguenza, non è possibile stabilire fin da subito dei valori di peso per i confronti a coppie già eseguiti. Possiamo però stabilire, per ogni attributo, dei livelli predefiniti:

- studi (S)
 - 1. ottimi (o)
 - 2. buoni (b)
 - 3. discreti (d)
- curriculum (C)
 - 1. eccellente (e)
 - 2. ottimo (o)
 - 3. buono (b)
- esperienza (E)
 - 1. molto significativa (ms)
 - 2. significativa (s)
 - 3. poco significativa (ps)
- conoscenza dell'inglese (I)
 - 1. professionale (p)
 - 2. scolastico (s)

Il fatto che non compaiano livelli di studi inferiori a discreto può indicare che tali livelli non sono considerati accettabili, per cui il candidato viene immediatamente scartato. In altre parole, tale soluzione non fa parte della regione ammissibile X e l'impatto associato ad esso non fa parte della regione degli impatti ammissibili F .

A questo punto, prima ancora di aver ascoltato un qualsiasi candidato, confrontiamo a coppie i livelli, e otteniamo un vettore di pesi relativi fra i livelli. Per gli studi:

$$\Lambda_S = \begin{array}{c|ccc} & o & b & d \\ \hline o & 1 & 3 & 6 \\ b & 1/3 & 1 & 2 \\ d & 1/6 & 1/2 & 1 \end{array} \Rightarrow u_S = \begin{bmatrix} 6/9 \\ 2/9 \\ 1/9 \end{bmatrix}$$

Si considera un livello di studi ottimo come moderatamente importante (peso 3 sulla scala di Saaty) rispetto a uno buono e qualcosa di mezzo fra fortemente e dimostrabilmente importante (peso 6) rispetto a una discreta.

Per il curriculum

$$\Lambda_C = \begin{array}{c|ccc} & e & o & b \\ \hline e & 1 & 4 & 6 \\ o & 1/4 & 1 & 3/2 \\ b & 1/6 & 2/3 & 1 \end{array} \Rightarrow u_C = \begin{bmatrix} 12/17 \\ 3/17 \\ 2/17 \end{bmatrix}$$

Per l'esperienza

$$\Lambda_E = \begin{array}{c|ccc} & ms & s & ps \\ \hline ms & 1 & 2 & 4 \\ s & 1/2 & 1 & 2 \\ ps & 1/4 & 1/2 & 1 \end{array} \Rightarrow u_E = \begin{bmatrix} 4/7 \\ 2/7 \\ 1/7 \end{bmatrix}$$

Per la conoscenza dell'inglese

$$\Lambda_I = \begin{array}{c|cc} & p & s \\ \hline p & 1 & 3 \\ s & 1/3 & 1 \end{array} \Rightarrow u_I = \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}$$

Occorre poi, come sempre, una matrice di confronti a coppie fra gli attributi, da cui si deduce un vettore dei pesi.

$$\Lambda_w = \begin{array}{c|cccc} & S & C & E & I \\ \hline S & 1 & 1 & 3 & 4 \\ C & 1 & 1 & 3 & 4 \\ E & 1/3 & 1/3 & 1 & 4/3 \\ I & 1/4 & 1/4 & 3/4 & 1 \end{array} \Rightarrow w = \begin{bmatrix} 12/31 \\ 12/31 \\ 4/31 \\ 3/31 \end{bmatrix}$$

A questo punto, un candidato con ottimi studi, curriculum eccellente, esperienza poco significativa e inglese scolastico riceve una valutazione pari a

$$u_S = 6/9 \quad u_C = 12/17 \quad u_E = 1/7 \quad u_I = 1/4$$

da cui, essendo $U = [u_S \ u_C \ u_E \ u_I]$

$$Uw = \begin{bmatrix} 6/9 & 12/17 & 1/7 & 1/4 \end{bmatrix} \cdot \begin{bmatrix} 12/31 \\ 12/31 \\ 4/31 \\ 3/31 \end{bmatrix} = 0.574$$

Invece, un candidato con buoni studi, ottimo curriculum, esperienza molto significativa e inglese professionale riceve una valutazione pari a

$$u_M = 2/9 \quad u_C = 3/17 \quad u_E = 4/7 \quad u_I = 3/4$$

da cui

$$Uw = \begin{bmatrix} 2/9 & 3/17 & 4/7 & 3/4 \end{bmatrix} \cdot \begin{bmatrix} 12/31 \\ 12/31 \\ 4/31 \\ 3/31 \end{bmatrix} = 0.301$$

ed è quindi considerato inferiore. Tutto questo vale a priori, prima ancora che i candidati si presentino.

7.2.9 Rank reversal nella teoria dell'utilità

Nella teoria dell'utilità classica il *rank reversal* è impossibile se si costruiscono correttamente le funzioni di utilità normalizzata. Tuttavia, il fenomeno può presentarsi se la costruzione è scorretta. Il caso più tipico è quello in cui in una fase successiva si aggiungono nuove soluzioni a X , il cui impatto cade fuori della regione F originale, modificando i valori estremi di alcuni attributi, e quindi falsandone la

normalizzazione². Di conseguenza, le altre alternative vedono cambiare la loro utilità normalizzata rispetto a tali attributi e le variazioni non sono uniformi. In particolare, se la nuova soluzione è molto buona per un attributo, le soluzioni che precedentemente erano migliori sembrano peggiorare molto, mentre quelle peggiori lo fanno di poco. È possibile che queste variazioni cambino l'ordinamento delle soluzioni. In teoria, i tassi marginali di sostituzione dovrebbero cambiare per tener conto della nuova scalatura, ma questo richiede che vengano ritrattati correttamente.

Esempio 59 Un problema ha due alternative A e B , con impatti $f(A) = (100, 0)$ e $f(B) = (90, 100)$, che rappresentano costi e hanno funzioni di utilità lineari. Le utilità normalizzate sono $U_A = (1, 0)$ e $U_B = (0, 1)$. Se i pesi delle due componenti della funzione di utilità sono $w_1 = 0.6$ e $w_2 = 0.4$, le utilità delle due alternative risultano:

$$u(A) = 0.6 \cdot 1 + 0.4 \cdot 0 = 0.6 \quad u(B) = 0.6 \cdot 0 + 0.4 \cdot 1 = 0.4$$

e vince A .

Se però compare una nuova alternativa C , con impatto $f(C) = (50, 0)$ e si rinormalizza l'utilità u_1 senza cambiare i pesi (e qui sta l'errore), U_A rimane lo stesso, mentre $U_B = (0.8, 1)$. Quindi l'utilità u_A rimane 0.6 e l'utilità u_B cresce da 0.4 a $u_B = 22/25 = 0.88$, diventando la migliore. In realtà, la modifica dell'intervallo di variazione di f_1 avrebbe dovuto condurre a rimodulare i pesi, come se l'unità di misura delle due componenti dell'utilità fosse cambiata.

7.3 Metodi ELECTRE

I metodi ELECTRE (ELimination Et Choix Traduisant la REalité³), sviluppati da B. Roy⁴ a partire dagli anni '60, partono da una critica di fondo all'ipotesi che il decisore sia capace di confrontare tutte le coppie di impatti.

Tipico a tale proposito è il *paradosso del caffè*, già brevemente presentato nella Sezione 3.5.5: due tazzine contengono la stessa quantità dello stesso caffè, ma la prima contiene una data quantità f di zucchero e la seconda un granello in più ($f' = f + \epsilon$). Per ϵ sufficientemente piccolo, le due tazze sono evidentemente indifferenti. Analogamente, la seconda tazza è indifferente a una terza tazza, che contiene una quantità $f'' = f + 2\epsilon$ di zucchero, e questa a una quarta che contiene una quantità $f + 3\epsilon$ di zucchero, ecc... Per la proprietà transitiva, due tazze poste all'estremo della catena e contenenti quantità f e $2f$ di zucchero sono anch'esse indifferenti. Ma questo è ovviamente falso. La teoria classica dell'utilità evita il paradosso assumendo che il decisore sappia discriminare ciascuno degli anelli della catena, indicando sempre quale sia migliore dell'altro.

Estendiamo il concetto a una situazione con molti attributi, per introdurre la soluzione proposta dai metodi ELECTRE. Supponiamo che la tazza di caffè A contenga un caffè di qualità superiore con la quantità ideale di zucchero più un granello, mentre la tazza di caffè B contiene un caffè di qualità media con la quantità ideale di zucchero. Secondo l'approccio di Pareto, le due tazze sono incomparabili. Secondo la teoria dell'utilità classica, sono comparabili a patto di stabilire un tasso

²Nuove soluzioni possono comparire durante il processo di modellazione a causa del fatto che tale processo non è lineare, ma procede per iterazioni successive. Per esempio, in una grande opera, le soluzioni vengono dettagliate e modificate per renderle più desiderabili via via che l'analisi procede e ne illustra le proprietà. Di solito, le nuove soluzioni non sono molto distanti dalla regione F originale, per cui l'effetto è trascurabile, ma in generale va tenuto presente.

³Eliminazione e scelta che traducono la realtà

⁴Bernard Roy (1934), matematico francese.

marginale di sostituzione fra gli attributi “qualità del caffè” e “quantità di zucchero”. I metodi ELECTRE seguono una via ancora diversa. Si parla di metodi ELECTRE, al plurale, perché nel tempo ne sono stati proposti molti, diversi tra loro. Qui presentiamo un riassunto dei concetti principali, senza fare riferimento esplicito all’uno o all’altro metodo.

7.3.1 La relazione di surclassamento

I metodi ELECTRE partono dalla relazione paretiana di preferenza, che considera A preferibile a B quando è non peggiore per tutti gli attributi, e la estendono ammettendo che A sia preferibile a B anche quando è peggiore per alcuni attributi, purché la differenza non superi una data soglia. Questo tipo particolare di relazione di preferenza è in genere definita *relazione di surclassamento* S (dal francese *surclasser*)⁵. Nel seguito, assumeremo che tutti gli indicatori costituiscano benefici per il decisore⁶. Inoltre, tratteremo problemi finiti, nei quali alternative e impatti si corrispondono biunivocamente, e quindi la distinzione fra i due concetti si può sfumare.

Definizione 30 *Dati due impatti $f, f' \in F \subseteq \mathbb{R}^P$, diremo che f surclassa f' ($f \preceq_S f'$) in base alle soglie $\epsilon_l \geq 0$ quando per tutti gli indicatori $l \in P$ l’impatto f non è eccessivamente peggiore di f' rispetto alle soglie:*

$$f \preceq_S f' \Leftrightarrow f_l \geq f'_l - \epsilon_l \text{ per ogni } l \in P$$

È facile vedere che con $\epsilon_l = 0$ per ogni $l \in P$, la relazione di surclassamento diventa quella di preferenza paretiana. Poiché di solito $\epsilon_l > 0$, la relazione di surclassamento è più ricca, cioè contiene un insieme più ampio di coppie. La preferenza paretiana, quindi, implica il surclassamento, ma non viceversa.

La relazione di surclassamento è chiaramente riflessiva, ma *in generale non gode della proprietà transitiva*, come illustrato dall’esempio seguente.

Esempio 60 *La Tabella 7.14 riporta le utilità di quattro impatti A, B, C e D rispetto a quattro attributi f_l .*

	f_1	f_2	f_3	f_4
A	0.50	0.50	0.50	0.60
B	0.45	0.45	0.45	0.70
C	0.40	0.40	0.40	0.80
D	0.35	0.35	0.35	0.90

Tabella 7.14: Matrice di valutazione (esempio)

Se supponiamo che sia $\epsilon_l = 0.05$ ($l = 1, \dots, 4$), gli impatti A e B differiscono fra loro meno della soglia ϵ_l per quanto riguarda i primi tre attributi, mentre B è chiaramente meglio di A per quanto riguarda il quarto. Quindi, B surclassa A ($B \preceq_S A$), ma non viceversa. Per lo stesso motivo, C surclassa B e D surclassa C . Sembra di poter concludere che D surclassi A , ma questo è falso, perché D è nettamente peggiore di A rispetto a tre dei quattro criteri; D e A sono in effetti incomparabili.

$$D \preceq_S C \preceq_S B \preceq_S A \text{ ma } D \not\preceq_S A, C \not\preceq_S A, B \not\preceq_S A, D \not\preceq_S B$$

⁵In inglese, si usa il termine *outranking*.

⁶Si noti che nel caso paretiano avevamo fatto l’ipotesi opposta.

Riassumendo, in questo caso la relazione di preferenza paretiana sarebbe completamente vuota e non fornirebbe informazioni utili. Invece, la relazione di surclassamento, tradotta in grafo orientato, corrisponderebbe a un cammino che va da D a C , a B , ad A . Questo non significa però che la soluzione D sia la migliore e la soluzione A la peggiore: in effetti, A è meglio di D per almeno tre indicatori. Mentre nel caso paretiano le soluzioni private di archi entranti dominano le altre, la relazione di surclassamento è più difficile da trattare.

7.3.2 Raffinamento della relazione di surclassamento

La definizione di surclassamento introdotta più sopra produce facilmente relazioni troppo povere e simili a quella paretiana quando le soglie ϵ_l sono molto piccole, oppure relazioni troppo ricche, in cui tutti gli impatti si surclassano a vicenda, quando le soglie sono grandi. In entrambi i casi, la relazione dà scarse indicazioni al decisore.

I metodi ELECTRE, quindi, propongono di combinare la definizione fondamentale con altre condizioni: la relazione di surclassamento include le coppie di impatti che verificano tutte le condizioni. Ognuna di tali condizioni sarà un rilassamento della preferenza paretiana, in modo che la preferenza paretiana le implichi sempre tutte. Di conseguenza, la preferenza paretiana è sempre condizione sufficiente per il surclassamento. Le condizioni consentono di tener conto di considerazioni aggiuntive, per esempio esprimendo l'importanza relativa degli indicatori attraverso opportuni pesi w_l .

Il decisore deve indicare sia le soglie ϵ_l sia gli altri parametri che regolano le condizioni di surclassamento. I valori che attribuisce loro influenzano il risultato del processo di decisione. I pesi w_l sono positivi e di somma unitaria come nella teoria dell'utilità a molti attributi e nell' AHP, ma non vengono utilizzati come coefficienti moltiplicativi in una combinazione convessa, bensì con un approccio che richiama quello delle votazioni: un impatto sarà considerato tanto più preferibile a un altro quanto maggiore è il peso degli indicatori rispetto ai quali è migliore.

Per ogni coppia di impatti (f, f') , è possibile definire i seguenti tre valori:

- la somma dei pesi per gli attributi rispetto ai quali f_l è meglio di f'_l

$$w_{ff'}^+ = \sum_{l \in P: f_l > f'_l} w_l$$

- la somma dei pesi per gli attributi rispetto ai quali f_l è indifferente a f'_l

$$w_{ff'}^\equiv = \sum_{l \in P: f_l = f'_l} w_l$$

- la somma dei pesi per gli attributi rispetto ai quali f_l è peggiore di f'_l

$$w_{ff'}^- = \sum_{l \in P: f_l < f'_l} w_l$$

Grazie all'ipotesi di normalizzazione vale la semplice proprietà:

$$w_{ff'}^+ + w_{ff'}^\equiv + w_{ff'}^- = \sum_{l \in P} w_l = 1 \text{ per ogni } f, f' \in F$$

Condizioni di surclassamento

Dato un insieme di condizioni, la relazione di surclassamento è data dalle coppie di impatti che le rispettano tutte, e quindi si ottiene intersecando le relazioni definite da ciascuna condizione. Alcune di queste condizioni sono le seguenti:

1. il *superamento di soglie di comparabilità*: l'impatto f è non troppo peggio dell'impatto f' per tutti gli attributi:

$$f \preceq_{S_\epsilon} f' \Leftrightarrow f_l \geq f'_l - \epsilon_l \text{ per ogni } l \in P$$

dove $\epsilon_l \geq 0$ per ogni $l \in P$. Questa è la condizione fondamentale introdotta prima. Si noti che per $\epsilon_l = 0$ questa relazione coincide con la preferenza paretiana, mentre per $\epsilon_l > 0$ si allarga a includere altre coppie di impatti, senza mai perdere quelle della preferenza paretiana.

2. la *concordanza*: un insieme di attributi di peso sufficiente suggerisce che f non sia peggiore di f' :

$$f \preceq_{S_c} f' \Leftrightarrow c_{ff'} = w_{ff'}^+ + w_{ff'}^- \geq \alpha_c \text{ con } \alpha_c \in [0; 1]$$

Per $\alpha_c = 1$, si ottiene ancora la preferenza paretiana, dato che $w_{ff'}^- \leq 1 - \alpha_c = 0$. Per gli altri valori di α_c , la relazione si arricchisce di altre coppie.

3. la *discordanza*: nessun attributo si oppone con entità eccessiva al fatto che f sia meglio di f' :

$$f \preceq_{S_d} f' \Leftrightarrow d_{ff'} = \begin{cases} \max \left(\frac{\max\limits_{l \in P: f_l < f'_l} (f'_l - f_l)}{\max\limits_{l \in P} |f_l - f'_l|}, 0 \right) & (\text{per } f \neq f') \\ 0 & (\text{per } f = f') \end{cases} \leq 1 - \alpha_d \text{ con } \alpha_d \in [0; 1]$$

Il numeratore misura la massima differenza fra i due impatti limitatamente agli indicatori rispetto ai quali f' è migliore, cioè $f'_l > f_l$. Il denominatore misura la massima differenza fra i due impatti in valore assoluto. Il rapporto varia fra 0, quando f non è mai peggio di f' , e 1, quando f è peggiore di f' per l'indicatore con la massima differenza. Questo indice esprime il rammarico che si proverebbe eliminando f' a favore di f . Per $\alpha_d = 1$, si ottiene ancora la preferenza paretiana⁷, per altri valori la relazione si arricchisce di altre coppie di impatti. Siccome si confrontano attributi diversi, è opportuno che essi siano prima normalizzati, in modo da rendere il confronto indipendente dalle unità di misura adottate.

La logica fondamentale dei metodi ELECTRE sta nel tenere in considerazione sia i valori degli indicatori f_l sia i valori dei loro pesi w_l , ma trattarli separatamente. Questo consente di evitare il procedimento della combinazione lineare fra pesi e indicatori, che ha senso solo sotto ipotesi piuttosto forti, cioè che gli indicatori siano trasformati in qualche modo in funzioni di utilità sommabili fra loro, che siano adimensionali ovvero normalizzati, ecc... Inoltre, consente di graduare l'importanza relativa di indicatori e pesi nello stabilire il risultato. Se si ha più fiducia nei valori stimati per i pesi, si può stringere la soglia di concordanza (o soglie analoghe definite sui pesi), in modo che le relative condizioni cancellino dalla relazione di surclassamento molte coppie. Se si ha più fiducia nei valori degli indicatori, si può

⁷Nei testi, generalmente si usa α_d al secondo membro della diseguaglianza. Qui si è preferito il complemento $1 - \alpha_d$ perché in questo modo tutti le condizioni corrispondono alla preferenza paretiana quando la soglia è pari a 1.

stringere invece la soglia di discordanza (o soglie analoghe definite sugli indicatori), in modo che siano le condizioni sugli indicatori a cancellare molte coppie dalla relazione di surclassamento. Questo consente di ottenere relazioni di surclassamento che dipendono più dai pesi che dagli indicatori o viceversa. Va detto che, in netto contrasto con questa logica, alcuni autori hanno proposto condizioni nelle quali pesi e indicatori si combinano (in particolare, esiste una definizione pesata di discordanza, nella quale le differenze fra i valori degli indicatori sono moltiplicate per i relativi pesi), ma non ci addentriamo in queste varianti.

Il valore dei parametri influenza pesantemente sul risultato finale. Questo richiede, se non c'è un forte consenso, un'analisi di sensitività che mostri la comparsa e scomparsa di preferenze al variare dei parametri. Se l'analisi suggerisce che i risultati dipendono poco dai valori dei parametri, la scelta è corretta. Per le soglie di concordanza e discordanza in particolare, è stato proposto di fissare le soglie in maniera adattiva, assegnando loro i valori medi dei coefficienti di concordanza e discordanza rispetto a tutte le coppie di impatti:

$$\alpha_c = \frac{\sum_{f,f' \in F} c_{ff'}}{|F|(|F|-1)} \quad \alpha_d = \frac{\sum_{f,f' \in F} d_{ff'}}{|F|(|F|-1)}$$

Questi valori hanno una buona probabilità di cancellare una frazione consistente, ma non eccessiva, delle coppie potenziali, producendo una relazione finale che non sia né troppo povera né troppo ricca⁸.

Esempio 61 Consideriamo un semplice esempio con tre soluzioni, valutate in base a quattro indicatori. La Tabella 7.15 riporta le loro prestazioni sotto forma di benefici adimensionali in una matrice di valutazione $\{f_{xl}\}$ e i pesi normalizzati degli indicatori in un vettore w .

f_{xl}	f_1	f_2	f_3	f_4
A_1	1	0.7	0.8	0.8
A_2	0.6	0.5	1	1
A_3	0.4	1	0.8	0.6

w	0.3	0.4	0.2	0.1
-----	-----	-----	-----	-----

Tabella 7.15: Matrice di valutazione per un'applicazione dei metodi ELECTRE

Supponiamo che le soglie di comparabilità siano ampie ($\epsilon_l = 0.5$): tutte e tre le soluzioni si surclassano a vicenda, con l'eccezione della coppia (a_3, a_1) . Consideriamo ora i pesi degli indicatori e costruiamo la matrice di concordanza C , cioè la matrice dei coefficienti $c_{ff'}$:

C	a_1	a_2	a_3
a_1	1	0.7	0.6
a_2	0.3	1	0.6
a_3	0.6	0.4	1

Fissiamo una soglia di concordanza pari ad $\alpha_c = 0.5$. La relazione S_c comprende tutte le coppie con concordanza $c_{ff'} \geq \alpha_c$, cioè

$$a_1 \preceq_{S_c} a_2, \quad a_2 \preceq_{S_c} a_3 \quad \text{e} \quad a_3 \sim_{S_c} a_1$$

⁸Però rendono le soglie dipendenti dagli impatti considerati, e quindi ripropongono l'insidioso problema del *rank reversal*: un'alternativa cattiva può abbassare le soglie, aggiungendo relazioni di surclassamento che possono cambiare il risultato del processo decisionale.

oltre ovviamente alle coppie riflessive $a_i \preceq_{S_c} a_i$ per ogni $a_i \in X$.

Infine, dai valori degli indicatori deriva la matrice di discordanza D , cioè la matrice dei coefficienti $d_{ff'}$:

D	a_1	a_2	a_3
a_1	0	0.5	0.5
a_2	1	0	1
a_3	1	0.8	0

Fissiamo una soglia di discordanza pari ad $\alpha_d = 0.5$. La relazione S_d comprende tutte le coppie con discordanza $d_{ff'} \leq 1 - \alpha_d$, cioè

$$a_1 \preceq_{S_c} a_2 \text{ e } a_1 \preceq_{S_c} a_3$$

oltre ovviamente alle coppie riflessive $a_i \preceq_{S_c} a_i$ per ogni $a_i \in X$.

Intersecando le tre relazioni, si ottiene la relazione finale di surclassamento

$$a_1 \preceq_{S_c} a_2$$

oltre alle coppie riflessive $a_i \preceq_{S_c} a_i$ per ogni $a_i \in X$.

7.3.3 Costruzione del nucleo

Se si considera un problema finito, gli impatti in F possono essere descritti come un grafo i cui archi rappresentano la relazione di preferenza. La regione paretiana corrisponde allora in tale grafo al sottoinsieme dei nodi di grado entrante nullo. La relazione di surclassamento, invece, non è transitiva, e questo impedisce di applicare lo stesso semplice criterio alla determinazione del sottoinsieme di soluzioni plausibili per la scelta. Nei metodi ELECTRE, questo sottoinsieme si ottiene con una procedura leggermente più complessa, di natura iterativa.

Definizione 31 Si definisce nucleo il sottoinsieme di soluzioni ottenuto con il seguente procedimento:

1. si parte con un nucleo vuoto;
2. si aggiunge al nucleo il sottoinsieme delle soluzioni non surclassate nel grafo corrente;
3. si elimina dal grafo ogni soluzione surclassata da una soluzione del nucleo;
4. se il grafo ridotto contiene solo il nucleo, si termina; altrimenti, si torna al punto 2.

Esempio 62 Dato il grafo di Figura 7.2, al primo passo G , D ed E entrano nel nucleo. Quindi, si estraggono dal grafo B (surclassata da E), C e F (surclassate da D ed E). Ora A entra nel nucleo. L'algoritmo termina.

La presenza di A nel nucleo può parere strana, dato che è surclassata da B e C che non ne fanno parte. Il motivo è che essere surclassati è più debole che essere dominati. Se le alternative valide sono quelle del nucleo e nessuna di loro surclassa A , significa che non ci sono motivi forti per eliminare A .

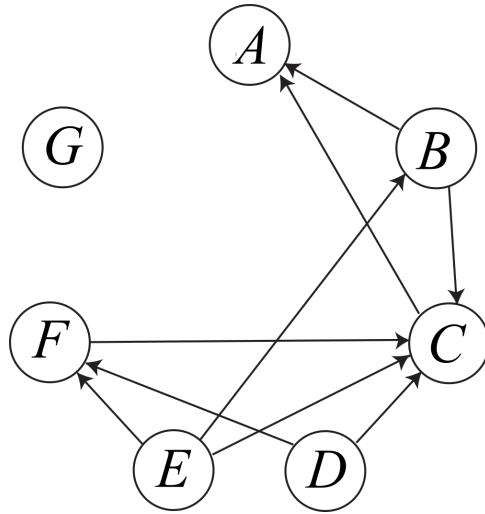


Figura 7.2: Un possibile grafo dei surclassamenti (gli autoanelli non sono riportati per semplicità)

Il caso dei grafi ciclici

Se il grafo dei surclassamenti contiene circuiti, cioè cicli orientati di surclassamenti, è possibile che la procedura non termini. Per esempio, potrebbero esistere coppie di soluzioni che si surclassano a vicenda, che nel grafo corrispondono a circuiti di due archi. A volte, per eliminare tali circuiti basta chiedere al decisore di scegliere fra le due alternative, oppure di rimodulare le soglie di comparabilità o i pesi, e quindi il grafo. In generale, però il problema riguarda circuiti più lunghi, e quindi più difficili da eliminare.

La seguente matrice valuta tre soluzioni rispetto a tre indicatori.

$f(x)$	f_1	f_2	f_3
A	3	2	1
B	2	1	3
C	1	3	2

Se consideriamo soglie $\epsilon_l = 2$ per $l = 1, \dots, 3$, i tre impatti si surclassano tutti a vicenda rispetto alle soglie di comparabilità. Se il peso dei tre indicatori è lo stesso, la condizione di concordanza può solo lasciare invariate tutte le coppie o cancellarle tutte dalla relazione, e lo stesso succede per la condizione di discordanza. La simmetria del problema impone di mantenere o cancellare tutte le coppie.

La letteratura ha proposto diversi approcci più o meno soddisfacenti per trattare il caso in cui non si può determinare un nucleo. Il più semplice è quello di accorpate ogni circuito in un supernodo, trattando i singoli nodi come se fossero indifferenti, dunque includendoli o escludendoli tutti insieme dal nucleo.

Analisi di sensitività sulle aree di surclassamento

È utile condurre un'analisi di sensitività rispetto alle soglie di concordanza α_c e di discordanza α_d , determinando per ogni soluzione i valori estremi che ne decretano il surclassamento da parte di un'altra soluzione⁹

⁹A rigore, l'analisi dovrebbe considerare solo le soluzioni del nucleo, perché solo in quel caso l'alternativa surclassata uscirebbe dal nucleo, ma i testi non sono chiari al proposito. Per semplicità

Definizione 32 Si dice area di dominanza di una soluzione l'insieme delle coppie (α_c, α_d) date le quali la soluzione risulta surclassata solo da sé stessa.

Osserviamo che abbassare le soglie aumenta il numero di coppie di impatti nella relazione di surclassamento, e quindi rende più facile che ciascuna soluzione risulti surclassata. Ogni arco potenziale (f', f) entrante in un nodo f ammette un valore minimo per α_c sotto il quale l'arco esiste e sopra il quale non esiste. Lo stesso vale per α_d . Di conseguenza, la regione del piano $\alpha_c \alpha_d$ nella quale un arco esiste è un rettangolo contenente l'origine. La regione del piano in cui un nodo è surclassato, cioè ha almeno un arco entrante, è l'unione di tali rettangoli. L'area di dominanza di una soluzione è il complemento di tali rettangoli.

Le soluzioni più forti sono difficilmente surclassabili, e quindi hanno aree di dominanza più ampie, che arrivano a lambire l'origine, mentre le soluzioni meno forti hanno aree inferiori, concentrate presso l'angolo opposto $(1, 1)$. Le aree di dominanza danno luogo a un ordinamento di soluzioni riflessivo e transitivo, ma non completo (in effetti, è un ordine parziale).

Esempio 63 Riprendiamo l'Esempio 61. La Figura DA FARE riporta i sei archi potenziali del grafo con le soglie relative ad α_c e α_d : PER ESEMPIO

Le aree di dominanza delle tre alternative sono descritte nella Figura DA FARE

Figura 7.3: Aree di dominanza delle tre alternative

7.3.4 Creazione di un ordinamento debole

Alcuni metodi ELECTRE prevedono una fase finale nella quale si introduce qualche criterio ulteriore per ordinare le soluzioni del nucleo. Come nel caso paretiano, si tratta di fare una parziale forzatura.

Ordinamento topologico

I nodi privi di archi entranti corrispondono a soluzioni non surclassate; tali soluzioni si propongono naturalmente come le migliori. Analogamente, i nodi privi di archi uscenti si propongono come i peggiori¹⁰.

Qualsiasi ordinamento topologico del grafo rispetta la relazione di surclassamento, nel senso che piazza i nodi surclassanti in posizioni precedenti a quelli surclassati. Si noti che l'ordinamento topologico richiede un grafo aciclico, per cui eventuali circuiti devono essere eliminati in partenza come già discusso nella Sezione 7.3.3. Tale grafo è parzialmente ordinato. L'ordinamento topologico introduce quindi una forzatura scegliendo uno dei diversi ordini totali compatibili con l'ordine parziale di partenza. Si noti che già solo partire dai nodi privi di archi entranti e procedere in avanti (*ordinamento crescente*) produce ordinamenti diversi dal partire da nodi privi di archi uscenti e procedere all'indietro (*ordinamento decrescente*). È stato anche proposto di determinare i due ordinamenti, assegnare a ogni nodo un indice di posizione nell'uno e nell'altro ordinamento e poi procedere una media dei due indici e usare tale media per ottenere l'ordinamento finale. Sono ovviamente metodi ampiamente empirici.

espositiva, consideriamo un generico surclassamento.

¹⁰Ovviamente, può succedere che un nodo ricada sotto entrambe le categorie.

Nell'esempio di Figura 7.2, l'ordinamento crescente parte da G . Quindi, siccome D ed E si surclassano a vicenda, vengono collassate in un solo nodo DE , e l'ordinamento può proseguire con DE , seguito da B e F , quindi C e infine A . L'ordinamento decrescente parte da A e G , poi C , la coppia B e F e ancora il macronodo DE . I due ordinamenti sono simili, ma la posizione del nodo isolato G è completamente diversa.

	A	B	C	D	E	F	G
Ordinamento crescente	7	4.5	6	2.5	2.5	4.5	1
Ordinamento decrescente	6.5	3.5	5	1.5	1.5	3.5	6.5
Composizione	6.75	4	5.5	2	2	3	3.75

La media delle posizioni dei nodi nei due ordinamenti suggerirebbe D ed E come soluzioni migliori, seguite da F , G , B , C per finire con A .

Ordinamento con indici aggregati

Si tratta di costruire per ogni impatto degli indici aggregati di concordanza e discordanza, a partire da quelli associati alle coppie di impatti:

- *indice di concordanza*: cerca di descrivere la soddisfazione legata alla scelta di un impatto; ha un valore alto quando l'impatto prevale sugli altri per molti attributi pesanti e gli altri prevalgono su di esso per pochi attributi leggeri:

$$C_f = \sum_{g \in F} (c_{fg} - c_{gf}) \quad f \in F$$

- *indice di discordanza*: cala quando il rammarico per una vittoria dell'impatto è piccolo, mentre il rammarico per una sconfitta è alto:

$$D_f = \sum_{g \in F} (d_{fg} - d_{gf}) \quad f \in F$$

Il *punto utopia*, cioè l'impatto che domina paretianamente tutti gli altri, e che in genere è irrealizzabile, avrebbe il massimo valore teorico di concordanza: $C^* = n-1$. Il *punto distopia*, cioè l'impatto dominato da tutti gli altri, avrebbe invece il minimo valore teorico: $C^\dagger = 1 - n$.

È stato proposto di ordinare gli impatti, e quindi le alternative, per indice di concordanza C_f decrescente o per indice di discordanza D_f decrescente. Ancora, è stato proposto di trattare i due indici come obiettivi e ricondursi a un problema di programmazione a due obiettivi. La forza dei due indici starebbe nel fatto che il primo si riferisce ai pesi, mentre il secondo ai valori degli attributi.

Esempio 64 Nell'Esempio 61, gli indici aggregati valgono: DA COMPLETARE

7.4 Esercizi* † ‡

Esercizio 1

Si consideri il problema decisionale con 5 alternative e 3 criteri, caratterizzato dai seguenti impatti (benefici) e pesi:

Attributi	a_1	a_2	a_3	a_4	a_5		pesi
f_1	30	0	60	90	120	w_1	0.1
f_2	5	6	4	2	7	w_2	0.6
f_3	18	26	17	20	16	w_3	0.3

Si normalizzi la matrice per ciascun criterio fra il valore minimo e il massimo, e si calcolino le matrici di concordanza e di discordanza.

Si costruisca il grafo dei surclassamenti in base alle tre condizioni seguenti:

1. concordanza con $\alpha_c = 0.75$, cioè $w_{ff'}^+ + w_{ff'}^- \geq 0.75$;
2. discordanza con $\alpha_d = 0.25$, cioè $\frac{\max_{l \in P: f_l < f'_l} |f_l - f'_l|}{\max_{l \in P} |f_l - f'_l|} \leq 0.75$

con soglie di comparabilità arbitrariamente ampie ($\epsilon_l = +\infty$ per ogni $l \in P$).

Si determini il nucleo del grafo dei surclassamenti.

Si ordinino le alternative in base all'indice di concordanza e in base all'indice di discordanza.

Soluzione

La matrice di valutazione normalizzata è:

Attributi	a_1	a_2	a_3	a_4	a_5
f_1	0.25	0.00	0.50	0.75	1.00
\tilde{f}_2	0.6	0.8	0.4	0.0	1.0
\tilde{f}_3	0.2	1.0	0.1	0.4	0.0

La matrice di concordanza è:

$$C = \begin{bmatrix} - & 0.1 & 0.9 & 0.6 & 0.3 \\ 0.9 & - & 0.9 & 0.9 & 0.3 \\ 0.1 & 0.1 & - & 0.6 & 0.3 \\ 0.4 & 0.1 & 0.4 & - & 0.3 \\ 0.7 & 0.7 & 0.7 & 0.7 & - \end{bmatrix}$$

La matrice di discordanza è¹¹:

$$D = \begin{bmatrix} - & 1.00 & 0.21 & 0.17 & 1.0 \\ 0.1 & - & 0.19 & 0.16 & 0.40 \\ 0.1 & 1.00 & - & 0.38 & 0.1 \\ 1.0 & 1.00 & 1.0 & - & 1.0 \\ 0.25 & 1.00 & 0.08 & 0.20 & - \end{bmatrix}$$

*Le soluzioni di questi esercizi non sono state ancora riviste. Eventuali segnalazioni di errore sono benvenute.

†Devo molti di questi esercizi a temi d'esame del prof. Alberto Colomni.

‡Benvenuti nei ruggenti anni '90. Pur avendoli vissuti da studente, sono io stesso incredulo al pensiero che un esercizio d'esame potesse richiedere tanti calcoli aritmetici.

¹¹Questa matrice era originariamente pesata, per cui ora è probabilmente errata.

Il grafo dei surclassamenti comprende 5 nodi corrispondenti alle alternative, e i seguenti archi: (a_1, a_3) , (a_2, a_1) , (a_2, a_3) .

Il nucleo è il sottoinsieme $N = \{a_2, a_4, a_5\}$.

Gli indici di concordanza delle alternative sono:

$$C_f = \sum_{f' \in F} c_{ff'} - \sum_{f' \in F} c_{f'f} = [-0.2 \ 2.0 \ -1.8 \ -1.6 \ 1.6]$$

da cui l'ordinamento $a_2 \prec a_5 \prec a_1 \prec a_4 \prec a_3$.

Gli indici di discordanza delle alternative sono¹²:

$$D_f = \sum_{f' \in F} d_{ff'} - \sum_{f' \in F} d_{f'f} = [0.03 \ -3.15 \ 1.90 \ 3.09 \ -1.87]$$

da cui l'ordinamento $a_2 \prec a_5 \prec a_1 \prec a_3 \prec a_4$.

Esercizio 2

Un centro sociale cerca una nuova sede: esistono quattro alternative (A, B, C e D) oltre all'alternativa 0 (restare nella sede attuale). Si è stabilito che la scelta tra le cinque alternative debba essere definitiva e che sarà fatta in base a tre fattori: costi, accessibilità e prestigio. È fornita una tabella indicante le utilità per ciascuna alternativa e fattore in una scala tra 0 e 100. È fornito anche un vettore di pesi dei tre fattori.

Indicatori	A	B	C	D	0		pesi
Costi	90	90	90	1	100	w_1	1/3
Accessibilità	12	13	10	100	37	w_2	1/3
Prestigio	30	1	5	100	10	w_3	1/3

Si rappresenti il problema per mezzo dell'analisi gerarchica, esprimendo le matrici dei confronti a coppie tra le alternative (rispetto a ciascun attributo) nella scala di Saaty. Nel calcolo dei valori, si scelga ogni volta il valore più vicino disponibile nella scala di Saaty.

Si dica se qualcuna delle matrici è coerente, indicando il vettore di ordinamento ad essa associato.

Si calcoli il valore del coefficiente di concordanza $c_{f_A f_D}$ definito dai metodi Electre.

Soluzione

Di questo esercizio non sono al momento disponibili le soluzioni.

Esercizio 3

Si consideri la seguente gerarchia in un problema di decisione a molti criteri:

- il decisore intende ottimizzare due criteri principali f_1 e f_2 , per i quali è nota la seguente matrice di confronti a coppie costruita con la scala di Saaty:

¹²Gli indici erano originariamente pesati, per cui ora sono probabilmente errati.

	f_1	f_2
f_1	1	4
f_2	1/4	1

- il criterio f_1 raccoglie due sottocriteri s_{11} e s_{12} , con la seguente matrice di confronti a coppie:

	s_{11}	s_{12}
s_{11}	1	1/9
s_{12}	9	1

- il criterio f_2 raccoglie due sottocriteri s_{21} e s_{22} , con la seguente matrice di confronti a coppie:

	s_{21}	s_{22}
s_{21}	1	3
s_{22}	1/3	1

- vi sono tre alternative a_1 , a_2 e a_3 ;
- sono noti i confronti a coppie fra le tre alternative secondo il sottocriterio s_{11} :

s_{11}	a_1	a_2	a_3
a_1	1	3	6
a_2	1/3	1	2
a_3	1/6	1/2	1

- sono noti i confronti a coppie fra le tre alternative secondo il sottocriterio s_{12} :

s_{11}	a_1	a_2	a_3
a_1	1	1/4	1/4
a_2	4	1	1
a_3	4	1	1

- sono noti i confronti a coppie fra le tre alternative secondo il sottocriterio s_{21} :

s_{11}	a_1	a_2	a_3
a_1	1	1/4	2
a_2	4	1	8
a_3	1/2	1/8	1

- sono noti i confronti a coppie fra le tre alternative secondo il sottocriterio s_{22} :

s_{11}	a_1	a_2	a_3
a_1	1	7	2
a_2	1/7	1	1/3
a_3	1/2	3	1

Si dica se le matrici sono tutte coerenti, motivando la risposta.

Se vi sono matrici incoerenti, le si renda coerenti modificando solo le coppie di valori λ_{13} e λ_{31} .

Una volta rese le matrici coerenti, si calcolino i corrispondenti vettori di pesi.

Si ricomponga la gerarchia fino ad ottenere l'ordinamento finale delle alternative, indicando i vari passaggi.

Si inserisca una nuova alternativa a_4 in maniera da provocare un *rank reversal*.

Soluzione

Le matrici sono tutte coerenti, tranne quella relativa ai confronti a coppie fra le alternative rispetto al sottocriterio s_{22} .

Si può rendere coerente tale matrice sostituendola con:

s_{11}	a_1	a_2	a_3
a_1	1	7	$7/3$
a_2	$1/7$	1	$1/3$
a_3	$3/7$	3	1

I vettori di pesi sono:

- per i criteri c_1 e c_2 : $[0.8 \ 0.2]^T$;
- per i sottocriteri s_{11} e s_{12} : $[0.1 \ 0.9]^T$;
- per i sottocriteri s_{21} e s_{22} : $[0.75 \ 0.25]^T$;
- per le alternative rispetto al criterio s_{11} : $[0.66 \ 0.22 \ 0.11]^T$;
- per le alternative rispetto al criterio s_{12} : $[0.11 \ 0.44 \ 0.44]^T$;
- per le alternative rispetto al criterio s_{21} : $[0.18 \ 0.72 \ 0.09]^T$;
- per le alternative rispetto al criterio s_{22} : $[0.63 \ 0.09 \ 0.27]^T$.

Se ne ricavano le seguenti pseudoutilità.

- $u_{a_1} = 0.053 + 0.079 + 0.027 + 0.031 = 0.190$;
- $u_{a_2} = 0.018 + 0.317 + 0.108 + 0.005 = 0.448$;
- $u_{a_3} = 0.009 + 0.317 + 0.013 + 0.014 = 0.362$.

che portano all'ordinamento: $a_2 \prec a_3 \prec a_1$.

Per produrre un rank reversal, è necessario inserire un'alternativa che si comporti “circa” come l'alternativa vincente a_2 nelle quattro matrici dei confronti a coppie fra alternative¹³.

Esercizio 4

È dato il seguente problema decisionale con tre alternative e quattro attributi (benefici), cui sono associati i pesi w_i ($i = 1, \dots, 4$):

Attributi	A	B	C		pesi
u_1	0	100	80	w_1	0.25
u_2	100	83	0	w_2	0.30
u_3	70	20	100	w_3	0.40
u_4	40	100	20	w_4	0.05

Si dica se esistono relazioni di surclassamento, quali e perché, con soglie di concordanza $\alpha_c = 0.70$ e di discordanza $\alpha = 0.60$.

¹³Credo che il motivo sia che in questo modo le pseudoutilità di a_2 approssimativamente si dimezzano, favorendo a_3 , ma non ho verificato.

Soluzione

La matrice di concordanza del problema è:

$$C = \begin{bmatrix} - & 0.7 & 0.35 \\ 0.3 & - & 0.6 \\ 0.65 & 0.4 & - \end{bmatrix}$$

La matrice di discordanza è¹⁴:

$$D = \begin{bmatrix} - & 1 & 0.8 \\ 0.5 & - & 1 \\ 1 & 1 & - \end{bmatrix}$$

Non c'è alcun surclassamento, perché i valori di discordanza sono tutti $> 1 - \alpha_d$. La matrice di concordanza da sola permetterebbe il surclassamento $A \prec_{S_c} B$.

Esercizio 5

Si determini il nucleo in base alle definizioni dei metodi Electre per un problema con sette alternative a_i ($i = 1, \dots, 7$) e il grafo dei surclassamenti dato dagli archi: (a_1, a_2) , (a_1, a_4) , (a_2, a_4) , (a_2, a_7) , (a_4, a_3) , (a_4, a_6) , (a_3, a_7) .

Soluzione

Il nucleo è un insieme di alternative tali che non esistono surclassamenti tra le alternative in esso contenute e ogni alternativa scartata è surclassata da almeno un'alternativa del nucleo.

In base a questa definizione, si può:

1. includere nel nucleo a_1 e a_5 ;
2. escludere dal nucleo a_2 e a_4 ;
3. includere nel nucleo a_3 e a_6 ;
4. escludere dal nucleo a_7

ottenendo $N = \{a_1, a_3, a_5, a_6\}$.

¹⁴Originariamente, la matrice era pesata, per cui questa potrebbe essere sbagliata.

Parte III

Modelli con scenari molteplici

Capitolo 8

Modellazione dell'incertezza

La *programmazione in condizioni di incertezza* affronta problemi di decisione nei quali occorre scegliere tra alternative il cui impatto dipende non solo dalle scelte del decisore, ma anche da fattori esterni che non possono essere previsti esattamente. Questi problemi sono caratterizzati dall'avere

- un solo decisore;
- un solo obiettivo;
- ambiente incerto: l'impatto della decisione dipende da variabili che non sono sotto il controllo del decisore e sul cui valore si hanno solo informazioni parziali.

Tali problemi si possono formulare come:

$$\begin{aligned} \min_{x \in X} f(x, \omega) \\ \omega \in \Omega \end{aligned}$$

dove

- X è l'insieme delle *alternative* o *soluzioni* x ;
- Ω è l'insieme degli *scenari* ω (in statistica, *spazio degli esiti*);
- $f(x, \omega)$ è l'*impatto* della soluzione x e dello scenario ω , ed è un numero reale;
- impatti più bassi sono preferiti a impatti più alti (oppure viceversa).

È importante osservare che il decisore sceglie l'alternativa x , ma non lo scenario ω , e che sceglie x prima che lo scenario ω sia svelato. Se x venisse scelta dopo, il problema si ridurrebbe ad una famiglia di problemi di Programmazione Matematica parametrizzata in ω , cioè a risolvere il problema

$$\min_{x \in X} f(x, \omega)$$

per ciascun singolo valore di $\omega \in \Omega$. La soluzione sarebbe allora una *strategia* $x^*(\omega)$.

Alcuni esempi tipici di problemi di decisione in condizioni di incertezza sono:

- finanza: scelta del portafoglio azionario (incertezza sul valore futuro delle azioni);

- marketing: sviluppo di nuovi prodotti e campagne promozionali (incertezza sulla risposta dei clienti);
- ricerca e sviluppo: opportunità di effettuare ricerche (incertezza sull'ammontare di costi e ricavi);
- gare di appalto: definizione del prezzo di un'offerta (incertezza sui costi e sull'entità delle offerte dei concorrenti);
- trivellazioni petrolifere: scelta di dove e quando effettuarle (incertezza sulle dimensioni del giacimento).

I problemi in condizioni di incertezza si dividono in due grandi categorie:

1. decisioni in condizioni di *ignoranza*: dello scenario ω si sa solo che cade in Ω ;
2. decisioni in condizioni di *rischio*: *si conosce la probabilità π_ω di ciascuno scenario $\omega \in \Omega$* (se Ω è un insieme discreto) *ovvero la densità di probabilità $\pi(\omega)$* (se Ω è un insieme continuo).

Per entrambe è possibile definire un concetto di dominanza fra soluzioni che consente di sfoltire l'insieme delle soluzioni interessanti.

8.1 Modellazione dell'incertezza

Vi sono due modi principali di descrivere situazioni incerte, che si riflettono in due forme principali di insiemi degli scenari Ω :

1. *descrizione a scenari*: Ω è un insieme finito, nel quale i singoli scenari sono elencati esplicitamente;
2. *descrizione a intervalli*: Ω è il prodotto cartesiano di un numero finito di intervalli, cioè ogni variabile esogena ω_k assume valori in un intervallo, indipendentemente dalle altre.

Questi non sono gli unici casi possibili: Ω potrebbe assumere forme anche molto più sofisticate. Tuttavia, si tratta di due casi particolari molto frequenti, e che quindi è interessante discutere più in dettaglio.

Esempio 65 Consideriamo un problema di cammino minimo fra due nodi di un grafo orientato. Supponiamo che l'incertezza sia modellata da due variabili esogene ω_1 e ω_2 , che rappresentano il tempo di percorrenza per due archi del grafo, che descrivono due tratti della rete stradale su cui il traffico è particolarmente incerto.

Una descrizione a scenari potrebbe consistere nell'elencare le possibili combinazioni dei valori delle due variabili. Per esempio, Ω potrebbe consistere in tre scenari, rappresentati dalle coppie $\omega^{(1)} = (2, 1)$, $\omega^{(2)} = (5, 1)$ e $\omega^{(3)} = (2, 6)$, che sono mostrate come quadrati rossi in Figura 8.1. I tre scenari corrispondono a situazioni in cui, rispettivamente, c'è poco traffico in entrambe le strade, molto traffico nella prima e poco nella seconda, oppure poco nella prima e molto nella seconda. Questo modello esclude la possibilità che vi sia molto traffico su entrambe le strade.

Una descrizione a intervalli, invece, potrebbe definire un intervallo di valori leciti per ciascuna delle due variabili ω_1 e ω_2 . La Figura 8.2 mostra con il tratteggio l'insieme degli scenari possibili Ω , che è il prodotto cartesiano degli intervalli $[2, 5]$ e $[1, 6]$ ed è un insieme continuo.

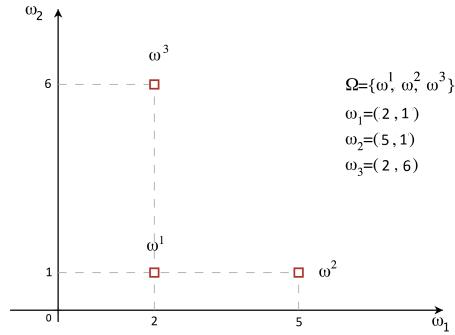


Figura 8.1: Descrizione a scenari dell'incertezza in un problema di cammino minimo

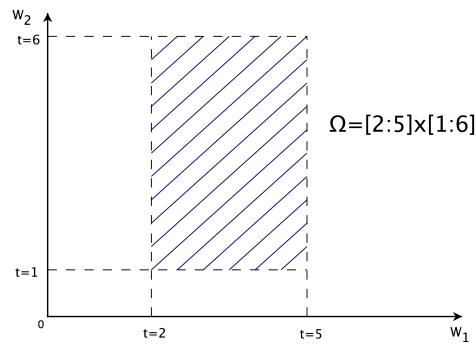


Figura 8.2: Descrizione a intervalli dell'incertezza in un problema di cammino minimo

La principale differenza fra le due descrizioni è che spesso nei modelli a intervalli lo scenario pessimo è comune a tutte le soluzioni (non sempre: dipende dal problema), mentre nei modelli a scenari in generale ogni soluzione ha il proprio scenario pessimo. Vedremo nella Sezione 9.9 che questo può facilitare il trattamento nei modelli a intervalli, con conseguenze importanti sulla complessità computazionale.

8.2 Dominanza forte

Definizione 33 Si dice che un'alternativa x domina fortemente un'alternativa x' quando il suo impatto è almeno altrettanto buono in tutti gli scenari $\omega \in \Omega$

$$x \preceq x' \Leftrightarrow f(x, \omega) \leq f(x', \omega) \text{ per ogni } \omega \in \Omega$$

Non sfuggirà la stretta analogia formale con il concetto di preferenza paretiana (si veda la Definizione 14 nella Sezione 5): è come se gli scenari fossero degli attributi distinti, rispetto a ciascuno dei quali si vorrebbe una prestazione ottimale. La differenza fondamentale è che gli scenari possono essere infiniti, e avere persino la potenza del continuo. La conclusione è però la stessa: le alternative strettamente dominate (cioè dominate, ma non dominanti) si possono scartare a priori; scartare le alternative strettamente dominate è l'unica cosa che si può fare razionalmente se non si aggiunge ulteriore informazione al problema.

Definizione 34 Nel caso finito, l'impatto $f(x, \omega)$ si può rappresentare con una matrice di valutazione U , che ha sulle righe le alternative x e sulle colonne gli scenari ω .

Questa matrice ha lo stesso nome di quella che rappresenta gli impatti in funzione di alternative x e attributi l nella Programmazione a Molti Attributi (vedi la Sezione 3.3), con gli scenari che prendono il posto degli attributi. Le due matrici non sono la stessa cosa, ma ci sono alcune analogie formali nel modo di trattarle. La prima è che per ricavare le alternative non dominate dalla matrice di valutazione si procede per confronti a coppie fra le righe della matrice, esattamente come nel caso paretiano (si veda la Sezione 5.3.1).

Esempio 66 Consideriamo una scelta di investimento su un dato orizzonte temporale. Vi sono quattro alternative:

1. fondo azionario
2. fondo obbligazionario
3. titoli di stato
4. titoli "tossici"

e un solo obiettivo, il rendimento, che deve essere massimizzato. Quindi, i valori migliori sono i più alti, contrariamente a quanto assunto nella Definizione 33: il concetto di dominanza forte va adattato, oppure bisogna cambiare segno alla funzione obiettivo. Il rendimento di ciascuna alternativa non è però noto a priori, dato che dipende anche dall'andamento dell'economia nel periodo considerato, che modelliamo con tre possibili scenari:

1. recessione
2. crescita moderata

Rendimento	Recessione	Crescita moderata	Crescita forte
Azioni	-25%	0%	35%
Obbligazioni	-10%	5%	15%
Titoli di stato	8%	8%	8%
Titoli tossici	-5%	6%	8%

Tabella 8.1: Matrice di valutazione per un problema di investimento con quattro alternative e tre scenari (l'impatto è il rendimento)

3. crescita forte

La Tabella 8.1 riporta la matrice di valutazione del problema di investimento. I titoli tossici sono dominati dai titoli di stato. Infatti, questi ultimi hanno rendimento strettamente migliore in caso di recessione e crescita moderata (8% contro -5% e 8% contro 6%) e uguale in caso di crescita forte (8%). Quindi i titoli "tossici" si possono razionalmente escludere dalla scelta, mentre le altre alternative sono tutte razionalmente difendibili.

8.3 Exercises*

Esercizio 1

Soluzione

*Le soluzioni di questi esercizi non sono state ancora riviste. Eventuali segnalazioni di errore sono benvenute.

Capitolo 9

Decisioni in condizioni di ignoranza

La condizione di ignoranza è la situazione più problematica per un decisore, dato che l'informazione sulla parte esogena del problema si riduce a sapere che cade in un dato insieme. Salvo il caso delle alternative dominate, non esistono criteri assoluti di scelta. Volendo ottimizzare l'obiettivo, bisognerebbe scegliere la coppia (x, ω) migliore, cioè fissare l'alternativa migliore e imporre lo scenario migliore. Nell'esempio dell'investimento finanziario, per massimizzare il rendimento, bisogna investire in azioni e procurare che l'economia cresca fortemente. Tuttavia, questo è fuori delle capacità di controllo del decisore, che può fissare solo la variabile x , e non la variabile ω . Il problema, posto in questi termini, non ha soluzione.

Gli approcci proposti in letteratura consistono nel ridefinire il problema, sostituendo a $f(x, \omega)$ una funzione ausiliaria $\phi_{\Omega}(x)$, che dipende da x e dall'intero insieme Ω , ma non dal valore (ignoto) di ω , e quindi ottimizzare quest'ultima.

Definizione 35 Si chiama criterio o logica di scelta ogni definizione di $\phi_{\Omega}(x)$ volta a sostituire l'impatto $f(x, \omega)$.

Passiamo ora in rassegna diverse logiche di scelta proposte in letteratura. Vedremo che ognuna di tali logiche ha qualche buon motivo a suo favore, ma nessuna soddisfa tutte le proprietà che sarebbero desiderabili per una scelta razionalmente fondata. La scelta fra loro dipende quindi solo dall'atteggiamento del decisore. Applicheremo ciascuna logica a un esempio con quattro alternative ($X = \{x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}\}$) e quattro scenari ($\Omega = \{\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)}\}$), la cui matrice di valutazione è riportata in Tabella 9.1 (si tratta di costi).

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	2	2	4	3
$x^{(2)}$	3	3	3	3
$x^{(3)}$	4	0	4	6
$x^{(4)}$	3	1	4	4

Tabella 9.1: Matrice di valutazione di esempio: costi associati alle quattro alternative e ai quattro scenari possibili

9.1 Criterio del caso pessimo

Questo criterio, detto anche *criterio del pessimismo* o *criterio di Wald*¹, consiste nell'essere pessimista e assumere che per ogni propria scelta x il futuro riservi lo scenario che comporta il costo massimo $\omega^\dagger(x) = \arg \max_{\omega \in \Omega} f(x, \omega)$. Questa ipotesi consente di ricavare l'impatto come funzione della sola variabile di decisione x (e dell'intero insieme Ω), cioè di ridurre il problema alla minimizzazione di $\phi_{\text{worst}}(x) = f(x, \omega^\dagger(x))$

$$\min_{x \in X} \phi_{\text{worst}}(x) = \min_{x \in X} \max_{\omega \in \Omega} f(x, \omega)$$

La Tabella 9.2 riporta per ciascuna alternativa il valore del criterio e lo scenario che gli dà luogo.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{worst}}(x)$	$\omega^\dagger(x)$
$x^{(1)}$	2	2	4	3	4	$\omega^{(3)}$
$x^{(2)}$	3	3	3	3	3	$\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)}$
$x^{(3)}$	4	0	4	6	6	$\omega^{(4)}$
$x^{(4)}$	3	1	4	4	4	$\omega^{(3)}, \omega^{(4)}$

Tabella 9.2: Applicazione del criterio del caso pessimo a un problema di esempio con quattro alternative e quattro scenari (minimizzazione di costi)

A questo punto, basta selezionare l'alternativa migliore, cioè di costo minimo. Il criterio del caso pessimo suggerisce di ordinare le alternative come segue: $x^{(2)} \prec x^{(1)} \sim x^{(4)} \prec x^{(3)}$.

9.2 Criterio del caso ottimo

Il criterio del caso ottimo, detto anche *criterio dell'ottimismo*, è complementare a quello del caso pessimo: consiste nell'assumere che, per ogni scelta x del deciso, il futuro risponda proponendo lo scenario migliore $\omega^*(x) = \arg \min_{\omega \in \Omega} f(x, \omega)$. Questa ipotesi consente di ricavare l'impatto come funzione della sola variabile x (e dell'intero insieme Ω), riducendo il problema alla minimizzazione della funzione $\phi_{\text{best}}(x) = f(x, \omega^*(x))$

$$\min_{x \in X} \phi_{\text{best}}(x) = \min_{x \in X} \min_{\omega \in \Omega} f(x, \omega)$$

La Tabella 9.3 riporta per ciascuna alternativa il valore del criterio e lo scenario che gli dà luogo. Poi si seleziona l'alternativa migliore, cioè di costo minimo. Il criterio dell'ottimismo suggerisce di ordinare le alternative come segue: $x^{(3)} \prec x^{(4)} \prec x^{(1)} \prec x^{(2)}$.

9.3 Criterio di Hurwicz

I primi due criteri appaiono piuttosto sbilanciati nella valutazione degli scenari. Per garantire un maggiore equilibrio nella scelta, Hurwicz² propose di costruire una

¹Abraham Wald (1902-1950), matematico ebreo ungherese, fuggito negli Stati Uniti dopo l'annessione dell'Austria alla Germania nazista.

²Leonid Hurwicz (1917-2008), economista ebreo polacco, bloccato all'estero dall'invasione nazista della Polonia, e poi rifugiato negli Stati Uniti.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{best}}(x)$	$\omega^*(x)$
$x^{(1)}$	2	2	4	3	2	$\omega^{(1)}, \omega^{(2)}$
$x^{(2)}$	3	3	3	3	3	$\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)}$
$x^{(3)}$	4	0	4	6	0	$\omega^{(2)}$
$x^{(4)}$	3	1	4	4	1	$\omega^{(2)}$

Tabella 9.3: Applicazione del criterio dell'ottimismo a un problema di esempio con quattro alternative e quattro scenari (minimizzazione di costi)

logica di scelta che ne facesse una combinazione convessa:

$$\min_{x \in X} \phi_{\text{Hurwicz}}(x) = \min_{x \in X} \left[\alpha \max_{\omega \in \Omega} f(x, \omega) + (1 - \alpha) \min_{\omega \in \Omega} f(x, \omega) \right]$$

dove si definisce *coefficiente di pessimismo*³ il coefficiente $\alpha \in [0; 1]$ che pesa l'impatto pessimo, mentre il valore complementare $1 - \alpha$ pesa quello ottimo. Tale coefficiente consente di modularre il peso dei due scenari estremi nella scelta: il valore $\alpha = 1$ corrisponde al criterio del pessimismo, mentre il valore $\alpha = 0$ corrisponde a quello dell'ottimismo.

La Tabella 9.4 riporta per ciascuna alternativa il valore del criterio nell'ipotesi che il coefficiente di pessimismo sia $\alpha = 0.6$, cioè che si sia leggermente più pessimisti che ottimisti. L'ordine risultante è $x^{(4)} \prec x^{(2)} \prec x^{(1)} \prec x^{(3)}$.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{Hurwicz}}(x) (\alpha = 0.6)$
$x^{(1)}$	2	2	4	3	$0.4 \cdot 2 + 0.6 \cdot 4 = 3.2$
$x^{(2)}$	3	3	3	3	$0.4 \cdot 3 + 0.6 \cdot 3 = 3$
$x^{(3)}$	4	0	4	6	$0.4 \cdot 0 + 0.6 \cdot 6 = 3.6$
$x^{(4)}$	3	1	4	4	$0.4 \cdot 1 + 0.6 \cdot 4 = 2.8$

Tabella 9.4: Applicazione del criterio di Hurwicz con parametro $\alpha = 0.6$ a un problema di esempio con quattro alternative e quattro scenari (minimizzazione di costi)

9.3.1 Taratura del coefficiente di pessimismo

Dato che la soluzione dipende dal valore del coefficiente di pessimismo α , è opportuno discutere in quale modo si possa tarare questo parametro. Un modo di procedere, analogo a quello usato per determinare i coefficienti di peso delle utilità a molti attributi, è individuare (eventualmente inventandole, se non esistono) una coppia di alternative indifferenti fra loro, imporre l'uguaglianza dei relativi criteri $\phi_{\text{Hurwicz}}(x)$ e risolvere l'equazione in α che ne deriva. Il modo più semplice di farlo è scegliere un'alternativa x e chiedere al decisore di indicare il suo *equivalente certo*, cioè un'alternativa y , che in generale non è una vera alternativa del problema, con impatto uniforme in tutti gli scenari e che sia indifferente se confrontata con x .

Nell'esempio che stiamo studiando sceglieremo l'alternativa $x^{(3)}$: essa ha costo 6 nel caso peggiore e 0 nel caso migliore. Se il decisore indica che un'alternativa y

³Va da sé che alcuni autori definiscono invece un coefficiente di ottimismo, che moltiplica il valore ottimo dell'impatto, e usano il valore complementare come moltiplicatore dell'impatto pessimo. Inoltre, se l'impatto è un guadagno anziché un costo, le operazioni di minimo e massimo vanno opportunamente modificate.

con costo uniforme pari a 4 sarebbe indifferente a $x^{(3)}$, possiamo concludere che $\phi_{\text{Hurwicz}}(x^{(3)}) = \phi_{\text{Hurwicz}}(y)$, cioè

$$\begin{aligned} \alpha \max_{\omega \in \Omega} f(x^{(3)}, \omega) + (1 - \alpha) \min_{\omega \in \Omega} f(x^{(3)}, \omega) &= \alpha \max_{\omega \in \Omega} f(y, \omega) + (1 - \alpha) \min_{\omega \in \Omega} f(y, \omega) \Rightarrow \\ \Rightarrow 6\alpha + 0(1 - \alpha) &= 4\alpha + 4(1 - \alpha) \Rightarrow \alpha = \frac{2}{3} \end{aligned}$$

Non è affatto ovvio che il decisore sia capace di indicare un equivalente certo, dato che sugli scenari supponiamo di non avere alcun'altra informazione salvo il fatto che esistono.

9.3.2 Sensitività rispetto al coefficiente di pessimismo

Data l'incertezza del coefficiente di pessimismo, è in genere opportuno condurre uno studio di sensitività che mostri se la soluzione scelta rimarrebbe la stessa in presenza di variazioni piccole del suo valore. Lo studio richiede di costruire per ogni alternativa $x \in X$ la funzione $\phi_{\text{Hurwicz}}(x)$, che dipende dal parametro $\alpha \in [0; 1]$:

- $u(x^{(1)}) = 4\alpha + 2(1 - \alpha) = 4\alpha + 2 - 2\alpha = 2\alpha + 2$
- $u(x^{(2)}) = 3\alpha + 3(1 - \alpha) = 3\alpha + 3 - 3\alpha = 3$
- $u(x^{(3)}) = 6\alpha + 0(1 - \alpha) = 6\alpha$
- $u(x^{(4)}) = 4\alpha + 1(1 - \alpha) = 4\alpha + 1 - 1\alpha = 3\alpha + 1$

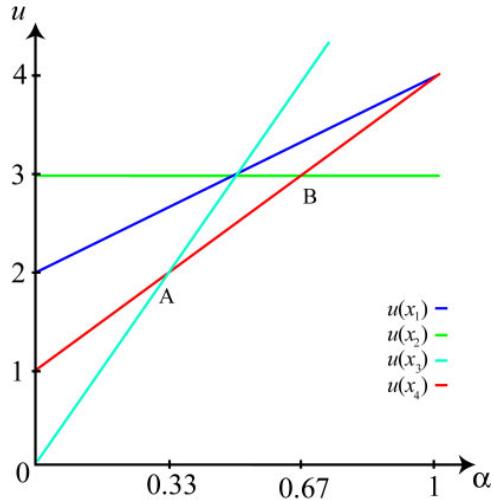


Figura 9.1: Studio di sensitività rispetto al coefficiente di pessimismo per un problema di esempio

La Figura 9.1 riporta il grafico delle quattro funzioni. Per ogni $\alpha \in [0; 1]$, si deve scegliere la soluzione, cioè la funzione, che minimizza il costo.

Dobbiamo quindi determinare i valori di α che individuano il punto A , intersezione di $\phi(x^{(3)})$ e $\phi(x^{(4)})$:

$$\phi(x^{(3)}) = \phi(x^{(4)}) \Rightarrow 6\alpha = 3\alpha + 1 \Rightarrow \alpha = \frac{1}{3} \Rightarrow A = \left(\frac{1}{3}, 2\right)$$

e il punto B , intersezione di $u(x^{(2)})$ e di $u(x^{(4)})$:

$$\phi(x^{(2)}) = \phi(x^{(4)}) \Rightarrow 3 = 3\alpha + 1 \Rightarrow \alpha = \frac{2}{3} \Rightarrow B = (\frac{2}{3}, 3)$$

Ne deriva che, se $0 \leq \alpha \leq \frac{1}{3}$ conviene l'alternativa $x^{(3)}$ (come con il criterio dell'ottimismo); se $\frac{1}{3} \leq \alpha \leq \frac{2}{3}$ conviene l'alternativa $x^{(4)}$; se $\frac{2}{3} \leq \alpha \leq 1$ conviene l'alternativa $x^{(2)}$ (come con il criterio del pessimismo). I valori estremi di α confermano ovviamente i risultati dei criteri dell'ottimismo e del pessimismo. Per valori intermedi, però, viene selezionata un'alternativa che i due criteri precedenti non avevano preso in considerazione.

Dal grafico in Figura 9.1 emerge anche che l'alternativa $x^{(1)}$ è sempre peggiore dell'alternativa $x^{(4)}$ (la funzione associata alla prima sta sopra a quella associata alla seconda). Questo non significa che $x^{(1)}$ sia dominata da $x^{(4)}$ (in effetti, è migliore nello scenario $\omega^{(3)}$): la situazione è simile a quella delle soluzioni paretiane, ma non supportate, che non risultavano mai ottimali rispetto a combinazioni lineari degli attributi (si veda la Sezione 5.6). Si tratta quindi di una limitazione intrinseca del criterio di Hurwicz: non è in grado di generare tutte le soluzioni non dominate giocando sul suo unico parametro. La limitazione è aggravata, rispetto al caso paretiano, dal fatto che il criterio considera solo i due scenari estremi, mentre il metodo dei pesi combinava tutti gli attributi del problema. D'altra parte, lo scopo del criterio di Hurwicz non è quello di generare le soluzioni, ma solo di ordinarle.

9.4 Criterio di equiprobabilità

Il criterio di equiprobabilità, detto anche *criterio di Laplace*, modifica quello di Hurwicz considerando tutti gli scenari, anziché solo quelli estremi. Non avendo informazioni sulla verosimiglianza degli scenari, ne combina gli impatti applicando a tutti lo stesso peso:

$$\min_{x \in X} \phi_{\text{Laplace}}(x) = \min_{x \in X} \frac{\sum_{\omega \in \Omega} f(x, \omega)}{|\Omega|}$$

Questa espressione presuppone che l'insieme degli scenari sia finito⁴. Si può interpretare il criterio di Laplace come la minimizzazione del costo atteso nell'ipotesi che tutti gli scenari abbiano la stessa probabilità di verificarsi (da cui il primo nome del criterio). Si ricordi però che supponiamo di non avere informazioni sulla probabilità degli scenari stessi.

La Tabella 9.5 riporta per ciascuna alternativa il valore del criterio, cioè il valor medio degli impatti nei vari scenari. L'ordine risultante è $x^{(1)} \prec x^{(2)} \sim x^{(4)} \prec x^{(3)}$.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{Laplace}}(x)$
$x^{(1)}$	2	2	4	3	$(2+2+4+3)/4 = 2.75$
$x^{(2)}$	3	3	3	3	$(3+3+3+3)/4 = 3.00$
$x^{(3)}$	4	0	4	6	$(4+0+4+6)/4 = 3.50$
$x^{(4)}$	3	1	4	4	$(3+1+4+4)/4 = 3.00$

Tabella 9.5: Applicazione del criterio di Laplace a un problema di esempio con quattro alternative e quattro scenari (minimizzazione di costi)

⁴Una generalizzazione a spazi campionari infiniti è possibile, ma va oltre i limiti di questo corso.

9.5 Criterio del rammarico

Questo criterio, detto anche *criterio del regret* o *criterio di Savage*⁵, consiste nel valutare il rammarico che il decisore proverebbe nel caso in cui si accorgesse di aver sbagliato scelta. Il criterio opera in più fasi:

1. si determina l'alternativa ottima per ogni scenario $x^*(\omega)$

$$x^*(\omega) = \arg \min_{x \in X} f(x, \omega)$$

2. si valuta il *rammarico* $\rho(x, \omega)$ (anche detto *regret*) associato a un'alternativa x e a uno scenario ω come la *perdita subita scegliendo x anziché l'alternativa ottima nello scenario ω* :

$$\rho(x, \omega) = f(x, \omega) - f(x^*(\omega), \omega) == f(x, \omega) - \min_{x \in X} f(x, \omega)$$

3. si valuta ogni soluzione con un criterio pari al rammarico massimo:

$$\phi_{\text{Savage}}(x) = \max_{\omega \in \Omega} \rho(x, \omega)$$

4. si sceglie la soluzione con il più piccolo rammarico massimo:

$$\min_{x \in X} \phi_{\text{Savage}}(x) = \min_{x \in X} \max_{\omega \in \Omega} \left[f(x, \omega) - \min_{x \in X} f(x, \omega) \right]$$

L'idea centrale è di ragionare in un'ottica di caso pessimo, come nel criterio di Wald, ma di definire il caso pessimo non come il caso in cui il costo è massimo, ma come quello in cui è massima la differenza fra ciò che si paga e ciò che si sarebbe potuto pagare scegliendo diversamente. In altre parole, questo criterio è relativo e opportunistico, anziché assoluto e conservativo: non punta a perdere il meno possibile, ma a restare indietro il meno possibile rispetto all'alternativa migliore.

La Tabella 9.6 riporta per ogni scenario l'alternativa ottima e il suo costo.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	2	2	4	3
$x^{(2)}$	3	3	3	3
$x^{(3)}$	4	0	4	6
$x^{(4)}$	3	1	4	4
$x^*(\omega)$	$x^{(1)}$	$x^{(3)}$	$x^{(2)}$	$x^{(2)}, x^{(4)}$
$f(x^*(\omega), \omega)$	2	0	3	3

Tabella 9.6: Applicazione del criterio di Savage a un problema di esempio con quattro alternative e quattro scenari (minimizzazione di costi): determinazione dell'alternativa ottima per ogni scenario

Da questa si può derivare la Tabella 9.7, che riporta il rammarico di ogni alternativa e scenario, cioè la differenza degli impatti dell'alternativa stessa rispetto alla migliore per quello scenario. L'ultima colonna della tabella riporta il rammarico nel caso peggiore per ogni alternativa. Ora cerchiamo di minimizzare tale rammarico: l'ordine risultante è $x^{(4)} \prec x^{(1)} \prec x^{(2)} \sim x^{(3)}$.

⁵Leonard Jimmie Savage, nato Leonard Ogashevitz (1917-1971), matematico e statistico americano, nato da ebrei russi fuggiti dalla Russia zarista.

$\rho(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{Savage}}(x)$
$x^{(1)}$	0	2	1	0	2 ($\omega^{(2)}$)
$x^{(2)}$	1	3	0	0	3 ($\omega^{(2)}$)
$x^{(3)}$	2	0	1	3	3 ($\omega^{(4)}$)
$x^{(4)}$	1	1	1	1	1 ($\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)}$)

Tabella 9.7: Applicazione del criterio di Savage a un problema di esempio con quattro alternative e quattro scenari (minimizzazione di costi): calcolo del rammarico per ogni alternativa e scenario e determinazione del rammarico massimo per ogni alternativa

9.6 Criterio delle eccedenze

Questo criterio, detto anche *criterio dei surpluses*, è complementare al criterio del rammarico: considera in ciascuno scenario l'eccedenza che il decisore ottiene rispetto all'alternativa pessima, e cerca di massimizzarla nel caso peggiore. Si procede quindi in modo analogo, ma complementare:

1. si determina l'alternativa pessima per ogni scenario $x^\dagger(\omega)$

$$x^\dagger(\omega) = \arg \max_{x \in X} f(x, \omega)$$

2. si valuta l'*eccedenza* $\sigma(x, \omega)$ (anche detta *surplus*) associata a un'alternativa x e a uno scenario ω come il *guadagno ottenuto scegliendo x anziché l'alternativa pessima nello scenario ω* :

$$\sigma(x, \omega) = f(x^\dagger(\omega), \omega) - f(x, \omega) = \max_{x \in X} f(x, \omega) - f(x, \omega)$$

3. si valuta ogni soluzione con un criterio pari all'eccedenza minima (caso pessimo):

$$\phi_{\text{surplus}}(x) = \min_{\omega \in \Omega} \sigma(x, \omega)$$

4. si sceglie la soluzione con la più grande eccedenza minima:

$$\max_{x \in X} \phi_{\text{surplus}}(x) = \max_{x \in X} \min_{\omega \in \Omega} \left[\max_{x \in X} f(x, \omega) - f(x, \omega) \right]$$

Rispetto agli altri criteri visti sinora, e in particolare al criterio del rammarico, la differenza sostanziale è che la funzione σ non misura una perdita rispetto alla scelta ottimale per ogni scenario, ma un “guadagno” rispetto alla scelta pessima per ogni scenario (nell'esempio, questo guadagno è in effetti una perdita minore). Quindi, lo scenario pessimo è quello in cui σ è minima, mentre lo scenario pessimo per il criterio del rammarico è quello che massimizza ρ . Come il criterio del rammarico, il criterio dell'eccedenza non usa un metro assoluto per i costi, ma un metro relativo a una situazione di riferimento: in questo caso, si cerca di stare il più possibile avanti rispetto all'alternativa peggiore.

La Tabella 9.8 mostra l'alternativa pessima per ogni scenario e il suo costo.

Quindi si calcola di quanto le altre alternative fanno risparmiare rispetto all'alternativa pessima. La Tabella 9.9 riporta tali valori per ogni alternativa e scenario. L'ultima colonna della tabella riporta l'eccedenza nel caso peggiore per ogni alternativa. I valori sono tutti nulli, a indicare che ogni soluzione ammette uno scenario rispetto alla quale è la peggiore. Ora dovremmo massimizzare tale eccedenza, ma il criterio in questo caso giudica le quattro soluzioni equivalenti: $x^{(1)} \sim x^{(2)} \sim x^{(3)} \sim x^{(4)}$.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^\dagger(\omega)$	$x^{(3)}$	$x^{(2)}$	$x^{(1)}, x^{(3)}, x^{(4)}$	$x^{(3)}$
$f(x^\dagger(\omega), \omega)$	4	3	4	6

Tabella 9.8: Applicazione del criterio delle eccedenze a un problema di esempio con quattro alternative e quattro scenari (minimizzazione di costi): determinazione dell'alternativa pessima per ogni scenario

$\sigma(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{surplus}}(x)$
$x^{(1)}$	2	1	0	3	0 ($\omega^{(3)}$)
$x^{(2)}$	1	0	1	3	0 ($\omega^{(2)}$)
$x^{(3)}$	0	3	0	0	0 ($\omega^{(1)}, \omega^{(3)}, \omega^{(4)}$)
$x^{(4)}$	1	2	0	2	0 ($\omega^{(3)}$)

Tabella 9.9: Applicazione del criterio delle eccedenze a un problema di esempio con quattro alternative e quattro scenari (minimizzazione di costi): calcolo dell'eccedenza per ogni alternativa e scenario e determinazione dell'eccedenza massima per ogni alternativa

9.7 Esempi

Consideriamo nel seguito una serie di esempi di applicazione dei sei criteri prima descritti, allo scopo di illustrare piccole varianti, come l'uso di valori parametrici per definire una strategia di decisione al variare di un dato incerto oppure il trattamento dell'incertezza nel caso continuo.

9.7.1 Installazione di un dispositivo

Una casa automobilistica deve decidere se installare su un nuovo modello un dispositivo di sicurezza del costo di 2500€ oppure non installarlo. Se il dispositivo non viene installato e si verifica un incidente, la casa automobilistica dovrà procedere all'installazione e ripagare le spese dell'incidente. Se installa il dispositivo, invece, dovrà pagare solo il costo dell'installazione, perché non sarà considerata responsabile di eventuali incidenti.

Il modello del problema prevede quindi due alternative:

- $x^{(1)}$: installare il dispositivo di sicurezza;
- $x^{(2)}$: non installarlo.

e due scenari:

- $\omega^{(1)}$: non si verifica alcun incidente;
- $\omega^{(2)}$: si verifica almeno un incidente.

A rigore, lo scenario $\omega^{(2)}$ si dovrebbe decomporre in un'intera gamma di sottoscenari, corrispondenti a incidenti di diversa gravità e con diversi costi. Per semplicità, però, fonderemo tutti questi sottoscenari in uno solo, introducendo un parametro C_I che misura il costo dell'incidente.

La Tabella 9.10 riporta la matrice di valutazione del problema: se la casa automobilistica installa il dispositivo, paga in entrambi gli scenari solo il costo fisso

di installazione di 2500 €; se non lo installa, nello scenario migliore non paga nulla, mentre nello scenario peggiore paga il dispositivo più il costo C_I dell'incidente.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$
$x^{(1)}$	2 500	2 500
$x^{(2)}$	0	$2 500 + C_I$

Tabella 9.10: Matrice di valutazione per un problema di scelta riguardo l'installazione di un dispositivo di sicurezza (minimizzazione di costi)

Ipotizziamo di essere in condizioni di completa ignoranza, cioè di non saper assegnare alcuna probabilità a $\omega^{(1)}$ e $\omega^{(2)}$.

Criterio del pessimismo La Tabella 9.11 riporta la valutazione del problema con il criterio del caso pessimo. Poiché $2 500 + C_I > 2 500$, per minimizzare il costo nel caso pessimo conviene adottare l'alternativa $x^{(1)}$, cioè installare il dispositivo di sicurezza.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{worst}}(x)$
$x^{(1)}$	2 500	2 500	2 500
$x^{(2)}$	0	$2 500 + C_I$	$2 500 + C_I$

Tabella 9.11: Applicazione del criterio del caso pessimo al problema di installazione

Criterio dell'ottimismo La Tabella 9.12 riporta la valutazione del problema con il criterio del caso ottimo. Questo consiglia di adottare l'alternativa $x^{(2)}$, cioè di non installare il dispositivo di sicurezza. Infatti, se non capitano incidenti, il costo da sostenere è nullo.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{best}}(x)$
$x^{(1)}$	2 500	2 500	2 500
$x^{(2)}$	0	$2 500 + C_I$	0

Tabella 9.12: Applicazione del criterio del caso ottimo al problema di installazione

Criterio di Hurwicz La Tabella 9.13 riporta la valutazione del problema con il criterio di Hurwicz: poiché il decisore non ha specificato alcun valore per il coefficiente di pessimismo α , esso compare come parametro nel criterio.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{Hurwicz}}(x)$
$x^{(1)}$	2 500	2 500	2 500
$x^{(2)}$	0	$2 500 + C_I$	$(2500 + C_I)(1 - \alpha)$

Tabella 9.13: Applicazione del criterio di Hurwicz con α parametrico al problema di installazione

Il costo dell'incidente C_I è un secondo parametro. La natura dei due parametri è molto diversa: α è in mano al decisore, mentre C_I è una variabile esogena generata

dall'ambiente. Tuttavia, si possono trattare entrambe come variabili rispetto a cui condurre uno studio di sensitività.

La Figura 9.2 riporta il grafico delle due funzioni $\phi_{\text{Hurwicz}}(x^{(1)})$ e $\phi_{\text{Hurwicz}}(x^{(2)})$ rispetto ad α e C_I : il piano orizzontale blu rappresenta l'alternativa $x^{(1)}$, la superficie curva rossa l'alternativa $x^{(2)}$.

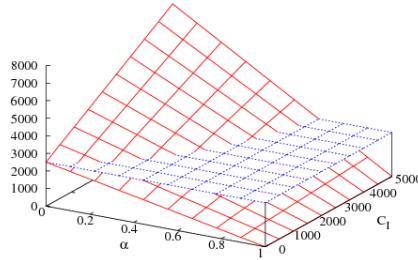


Figura 9.2: Studio di sensitività rispetto ad α e C_I per il problema di installazione: il piano orizzontale blu corrisponde all'installazione, la superficie curva rossa alla non installazione

Poiché $x^{(1)} \preceq x^{(2)}$ quando $\phi_{\text{Hurwicz}}(x^{(1)}) \leq \phi_{\text{Hurwicz}}(x^{(2)})$, dalla figura si ricava che

$$x^{(1)} \preceq x^{(2)} \Leftrightarrow 2500 < (1 - \alpha)(2500 + C_I) \Leftrightarrow C_I > \frac{\alpha}{1 - \alpha} 2500$$

La Figura 9.3 mostra la proiezione sul piano αC_I del grafico in Figura 9.2: la zona colorata in blu contiene le coppie (α, C_I) in cui conviene adottare l'alternativa $x^{(1)}$, quella in rosso le coppie in cui conviene l'alternativa $x^{(2)}$; le due aree sono divise dalla curva $C_I = 2500\alpha / (1 - \alpha)$, che tende all'infinito per $\alpha \rightarrow 0$ e poi va calando fino a zero.

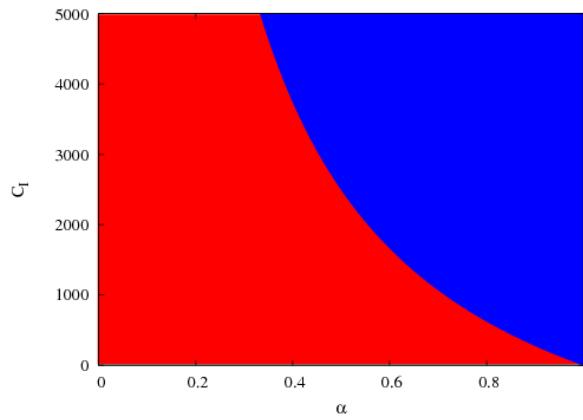


Figura 9.3: Studio di sensitività del problema di installazione rispetto ad α e C_I : supporto delle due alternative nello spazio dei parametri

In termini qualitativi:

- se il decisore ha un'alta propensione al rischio (valori bassi di α), gli conviene adottare la strategia $x^{(1)}$, cioè non installare il dispositivo, anche per valori alti (al limite, per qualsiasi valore) del costo di un eventuale incidente;
- se il decisore ha una bassa propensione al rischio (alti valori di α) gli conviene adottare la strategia $x^{(2)}$, cioè installare il dispositivo anche per stime basse del costo di un eventuale incidente;
- il valore di soglia del costo C_I che divide le due soluzioni cresce più che linearmente al crescere della propensione al rischio.

Criterio di Laplace Si assume che la probabilità che accada un incidente sia uguale a quella che non accada alcun incidente. La Tabella 9.14 riporta i corrispondenti valori del criterio di Laplace. Se ne deduce che $x^{(1)} \preceq x^{(2)}$ quando $2500 \leq 1250 + C_I/2$, cioè $C_I \geq 2500$.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{Laplace}}(x)$
$x^{(1)}$	2500	2500	2500
$x^{(2)}$	0	$2500 + C_I$	$1250 + C_I/2$

Tabella 9.14: Applicazione del criterio di Laplace al problema di installazione

Criterio del rammarico Il costo minimo e la corrispondente alternativa ottima per ciascuno dei due scenari sono riportati nell'ultima riga della Tabella 9.15.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	
$x^{(1)}$	2500	2500	
$x^{(2)}$	0	$2500 + C_I$	
$x^*(\omega)$	$x^{(2)}$	$x^{(1)}$	
$f(x^*(\omega), \omega)$	0	2500	

Tabella 9.15: Applicazione del criterio del rammarico al problema di installazione: determinazione delle soluzioni ottime per ogni scenario

La funzione di rammarico $\rho(x, \omega)$ assume i valori riportati nella Tabella 9.16; l'ultima colonna riporta il rammarico nel caso pessimo per ogni soluzione.

$\rho(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{regret}}(x)$
$x^{(1)}$	2500	0	2500
$x^{(2)}$	0	C_I	C_I

Tabella 9.16: Applicazione del criterio del rammarico al problema di installazione: calcolo della funzione rammarico e determinazione del caso pessimo

Anche in questo caso la soluzione non è univoca ma dipende dal costo dell'incidente. Poiché vogliamo minimizzare il rammarico nel caso peggiore, diremo che $x^{(1)} \preceq x^{(2)}$ quando $C_I \geq 2500$.

Criterio delle eccedenze L'alternativa pessima e il relativo costo per ciascuno dei due scenari sono riportati nell'ultima riga della Tabella 9.17.

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$
$x^{(1)}$	2 500	2 500
$x^{(2)}$	0	$2 500 + C_I$
$x^{\dagger}(\omega)$	$x^{(1)}$	$x^{(2)}$
$f(x^{\dagger}(\omega), \omega)$	$2 500 + C_I$	2 500

Tabella 9.17: Applicazione del criterio dell'eccedenza al problema di installazione: determinazione delle soluzioni pessime per ogni scenario

La funzione di eccedenza $\sigma(x, \omega)$ assume i valori riportati nella Tabella 9.18; l'ultima colonna riporta l'eccedenza nel caso pessimo per ogni soluzione. Si dovrebbe scegliere la soluzione che massimizza l'eccedenza nel caso pessimo, ma in questo caso le due soluzioni sono equivalenti da questo punto di vista: $x^{(1)} \sim x^{(2)}$.

$\sigma(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{surplus}}(x)$
$x^{(1)}$	0	C_I	0
$x^{(2)}$	2 500	0	0

Tabella 9.18: Applicazione del criterio dell'eccedenza al problema di installazione: calcolo della funzione eccedenza e determinazione del caso pessimo

9.7.2 Un investimento finanziario

Applichiamo ciascuna delle logiche su elencate all'esempio di investimento introdotto al principio del capitolo. In questo esempio gli impatti sono rendimenti, e vanno quindi massimizzati.

Criterio del caso pessimo Questo criterio dà luogo alla Tabella 9.19, secondo la quale si dovrebbe investire in titoli di stato.

$f(x, \omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{\text{worst}}(x)$	$\omega_{\text{worst}}(x)$
Azioni	-25%	0%	35%	-25%	Recessione
Obbligazioni	-10%	5%	15%	-10%	Recessione
Titoli di stato	8%	8%	8%	8%	Recessione

Tabella 9.19: Applicazione del criterio di Wald a un problema di investimento (massimizzazione di rendimenti)

Criterio del caso ottimo Questo criterio dà luogo alla Tabella 9.20, secondo la quale si dovrebbe investire in fondi azionari.

Criterio di Hurwicz Il criterio di Hurwicz con $\alpha = 0.6$, cioè con una visione bilanciata, ma leggermente più pessimista che ottimista, dà luogo alla Tabella 9.21, secondo la quale si dovrebbe investire in titoli di stato⁶.

⁶Poiché l'impatto è un rendimento, il parametro di pessimismo $\alpha = 0.6$ va applicato ai valori peggiori, cioè ai più bassi.

$f(x, \omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{\text{worst}}(x)$	$\omega_{\text{worst}}(x)$
Azioni	-25%	0%	35%	35	Crescita forte
Obbligazioni	-10%	5%	15%	15	Crescita forte
Titoli di stato	8%	8%	8%	8	Crescita forte

Tabella 9.20: Applicazione del criterio dell'ottimismo a un problema di investimento (massimizzazione di rendimenti)

$f(x, \omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{\text{Hurwicz}}(x)$
Azioni	-25%	0%	35%	$-25\% \cdot 0.6 + 35\% \cdot 0.4 = -1\%$
Obbligazioni	-10%	5%	15%	$-10\% \cdot 0.6 + 15\% \cdot 0.4 = 0\%$
Titoli di stato	8%	8%	8%	$+8\% \cdot 0.6 + 8\% \cdot 0.4 = 8\%$

Tabella 9.21: Applicazione del criterio di Hurwicz con $\alpha = 0.6$ a un problema di investimento (massimizzazione di rendimenti)

Studio di sensitività rispetto al coefficiente di pessimismo Tornando sul criterio di Hurwicz, si può approfondire il discorso conducendo anche l'analisi di sensitività. Definiamo per ogni investimento possibile il rendimento, funzione del parametro α :

$$\begin{aligned}\phi_A(\alpha) &= -25\alpha + 35(1-\alpha) &= 35 - 60\alpha \\ \phi_O(\alpha) &= -10\alpha + 15(1-\alpha) &= 15 - 25\alpha \\ \phi_T(\alpha) &= 8\alpha + 8(1-\alpha) &= 8\end{aligned}$$

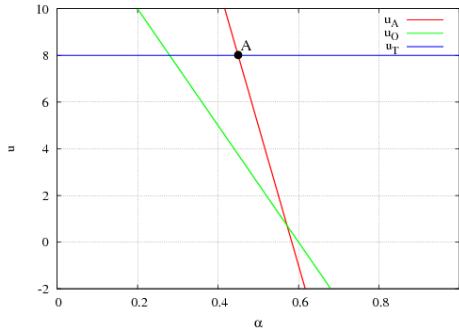


Figura 9.4: Studio di sensitività rispetto al coefficiente di pessimismo per il problema di investimento (massimizzazione di rendimenti)

La Figura 9.4 mostra i grafici delle tre funzioni. Per ogni $\alpha \in [0; 1]$, bisogna scegliere la funzione con il valore massimo perché stiamo massimizzando un rendimento. Anche senza fare calcoli, si nota che le obbligazioni non sono mai convenienti, poiché la relativa retta è sempre superata da almeno una delle altre due. Questo non significa che esse siano una soluzione dominata: il fenomeno è analogo a quello delle soluzioni paretiane non supportate (la combinazione lineare non è sufficiente a far emergere soluzioni dominate dalle combinazioni convesse delle altre). Il fenomeno è poi ulteriormente aggravato dal fatto che il criterio di Hurwicz combina solo due degli scenari, mentre il metodo dei pesi combinava tutti gli indicatori. Per individuare l'investimento migliore in funzione di α , bisogna

trovare il punto di intersezione A fra la retta delle azioni e quella dei titoli di stato: $\phi_A(\alpha) = \phi_T(\alpha) \Rightarrow 35 - 60\alpha = 8 \Rightarrow \alpha = 0.45$. Quindi, lo studio di sensitività conferma i risultati ottenuti con i metodi del pessimismo e dell'ottimismo: se si è ottimisti ($0 \leq \alpha \leq 0.45$), conviene investire in azioni; se si è pessimisti ($0.45 \leq \alpha \leq 1$), conviene investire in titoli di stato.

Criterio di Laplace Il criterio di Laplace dà luogo alla Tabella 9.22, che suggerisce, ancora una volta, di investire in titoli di stato.

$f(x, \omega)$				$\phi_{\text{Laplace}}(x)$
	Recessione	Crescita moderata	Crescita forte	
Azioni	-25%	0%	35%	$(-25\% + 0\% + 35\%)/3 = 3.33\%$
Obbligazioni	-10%	5%	15%	$(-10\% + 5\% + 15\%)/3 = 3.33\%$
Titoli di stato	8%	8%	8%	$(8\% + 8\% + 8\%)/3 = 8\%$

Tabella 9.22: Applicazione del criterio di Laplace a un problema di investimento (massimizzazione dei rendimenti)

Criterio del rammarico Il criterio del rammarico richiede anzi tutto di calcolare l'alternativa ottima per ogni scenario e il suo costo (vedi Tabella 9.23).

$f(x, \omega)$				
	Recessione	Crescita moderata	Crescita forte	
Azioni	-25%	0%	35%	
Obbligazioni	-10%	5%	15%	
Titoli di stato	8%	8%	8%	
$x^*(\omega)$	Titoli di stato	Titoli di stato	Azioni	
$f(x^*(\omega), \omega)$	8%	8%	35%	

Tabella 9.23: Applicazione del criterio di Savage a un problema di investimento (massimizzazione dei rendimenti): calcolo dell'alternativa ottima per ogni scenario

Quindi, richiede di valutare la funzione rammarico per ciascuna alternativa e scenario e trovarne il massimo per ciascuna alternativa (vedi Tabella 9.24). Investendo in azioni, la cosa peggiore che possa succedere è di andare in recessione, perdendo il 25% anziché guadagnare l'8%, con un rammarico del 33%. Investendo in obbligazioni, lo scenario peggiore è la crescita forte, che fa guadagnare il 15% anziché il 35%, con un rammarico del 20%; investendo in titoli di stato, lo scenario peggiore è la crescita forte, che fa guadagnare l'8% anziché il 35%, con un rammarico del 27%⁷.

Infine, si può procedere alla scelta dell'alternativa che produce il rammarico minimo, cioè investire in obbligazioni.

Criterio delle eccedenze Il criterio delle eccedenze dà anzi tutto luogo alla Tabella 9.25, che riporta l'alternativa pessima per ogni scenario e il suo costo.

Quindi, il criterio genera la Tabella 9.26, che riporta l'eccedenza per ciascuna alternativa e scenario, e nell'ultima colonna l'eccedenza minima per ogni alternativa. Investendo in azioni, la cosa peggiore che possa succedere è che vi sia recessione o crescita moderata, nel qual caso l'eccedenza è nulla; investendo in obbligazioni, lo

⁷Tutte queste percentuali sono riferite all'ammontare dell'investimento iniziale.

$\rho(x, \omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{Savage}(x)$
Azioni	$8\% - (-25\%) = 33\%$	$8\% - 0\% = 8\%$	$35\% - 35\% = 0\%$	33%
Obbligazioni	$8\% - (-10\%) = 18\%$	$8\% - 5\% = 3\%$	$35\% - 15\% = 20\%$	20%
Titoli di stato	$8\% - 8\% = 0\%$	$8\% - 8\% = 0\%$	$35\% - 8\% = 27\%$	27%

Tabella 9.24: Applicazione del criterio di Savage a un problema di investimento (minimizzazione dei rendimenti): calcolo del rammarico per ogni alternativa e scenario

$f(x, \omega)$	Recessione	Crescita moderata	Crescita forte
Azioni	-25%	0%	35%
Obbligazioni	-10%	5%	15%
Titoli di stato	8%	8%	8%
$x^\dagger(\omega)$	Azioni	Azioni	Titoli di stato
$f(x^\dagger(\omega), \omega)$	-25%	0%	8%

Tabella 9.25: Applicazione del criterio delle eccedenze a un problema di investimento (massimizzazione dei rendimenti): calcolo dell'alternativa pessima per ogni scenario

scenario pessimo è la crescita moderata, che fa guadagnare il 5% anziché lo 0%, con un'eccedenza del 5%; investendo in titoli di stato, lo scenario pessimo è la crescita forte, che fa guadagnare l'8%, con un'eccedenza nulla.

$\sigma(x, \omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{surplus}(x)$
Azioni	$-25\% - (-25\%) = 0\%$	$0\% - 0\% = 0\%$	$35\% - 8\% = 27\%$	0%
Obbligazioni	$-10\% - (-25\%) = 15\%$	$5\% - 0\% = 5\%$	$15\% - 8\% = 7\%$	5%
Titoli di stato	$8\% - (-25\%) = 33\%$	$8\% - 0\% = 8\%$	$8\% - 8\% = 0\%$	0%

Tabella 9.26: Applicazione del criterio delle eccedenze a un problema di investimento (minimizzazione dei rendimenti): calcolo dell'eccedenza per ogni alternativa e scenario

Infine, si può procedere alla scelta dell'alternativa che produce l'eccedenza massima: investire in obbligazioni.

9.7.3 Un esempio continuo*

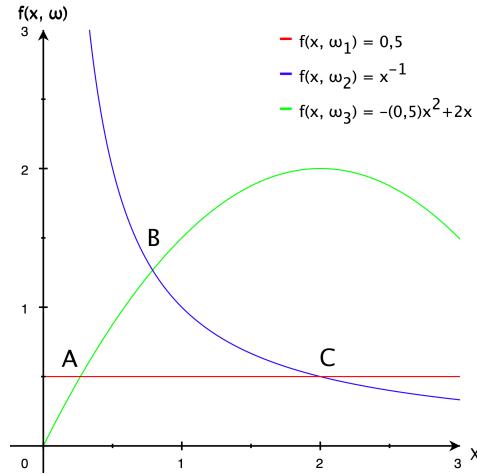
Consideriamo un problema nel quale la regione ammissibile X è un insieme continuo (un intervallo), mentre l'insieme degli scenari Ω rimane finito:

$$X = [10^{-9}, 3] \quad \Omega = \{\omega^{(1)}, \omega^{(2)}, \omega^{(3)}\}$$

Poichè X è continuo non si può descrivere la funzione impatto $f(x, \omega)$ con una tabella. Si può però descriverla, come nella Tabella 9.27, specificando una funzione $f(x, \omega_i)$ per ogni $\omega_i \in \Omega$. Supponiamo che gli impatti rappresentino dei costi da minimizzare.

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
$f(x, \omega)$	$\frac{1}{2}$	$\frac{1}{x}$	$-\frac{1}{2}x^2 + 2x$

Tabella 9.27: Valori dell'impatto $f(x, \omega)$ in ciascuno scenario $\omega_i \in \Omega$ Figura 9.5: Andamento dell'impatto $f(x, \omega)$ in ciascuno scenario $\omega_i \in \Omega$

La Figura 9.5 mostra i grafici delle tre funzioni impatto. Poiché saranno utili in seguito, riportiamo anche le coordinate dei punti di intersezione delle tre curve:

$$A = \left(2 - \sqrt{3}, \frac{1}{2}\right) \quad B = (0.79, 1.27) \quad C = \left(2, \frac{1}{2}\right)$$

Criterio del caso pessimo Per ogni $x \in X$ bisogna selezionare la funzione con valore massimo; della curva ottenuta bisogna selezionare il punto minimo. La Figura 9.6 mostra la soluzione suggerita dal criterio, che è $x_{\text{worst}}^* = x_B = 0.79$.

Criterio del caso ottimo Per ogni $x \in X$ bisogna selezionare la funzione con valore minimo; della curva ottenuta bisogna selezionare il punto minimo. La Figura 9.7 mostra la soluzione suggerita dal criterio, che è in $x_{\text{best}}^* = 10^{-9}$.

Criterio di Hurwicz e studio di sensitività La funzione che esprime il criterio di Hurwicz è:

$$\phi_{\Omega}(x) = \alpha \min[f(x, \omega^{(1)}), f(x, \omega^{(2)}), f(x, \omega^{(3)})] + (1-\alpha) \max[f(x, \omega^{(1)}), f(x, \omega^{(2)}), f(x, \omega^{(3)})]$$

La Figura 9.8 mostra l'applicazione del criterio di Hurwicz con diversi valori di α : le curve estreme rappresentano il criterio dell'ottimismo ($\alpha = 1$) e del pessimismo ($\alpha = 0$). Per ogni valore di α campionato viene segnata in neretto la soluzione del criterio, cioè il punto di minimo della curva. Per alcuni valori di α si trova una soluzione ($x_{\text{Hurwicz}}^* = 3$) diversa da quelle trovate con i criteri del pessimismo e dell'ottimismo.

La Tabella 9.28 riporta le soluzioni per ogni valore campionario di α .

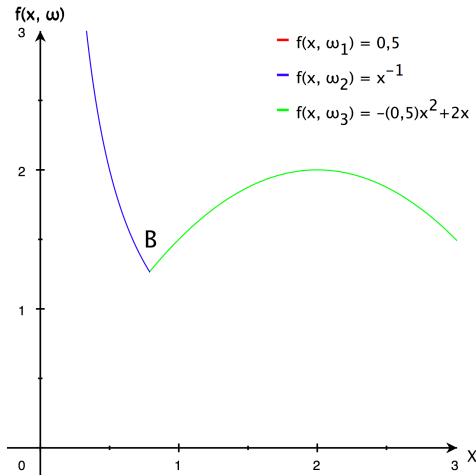


Figura 9.6: Applicazione del criterio del caso pessimo all'esempio continuo

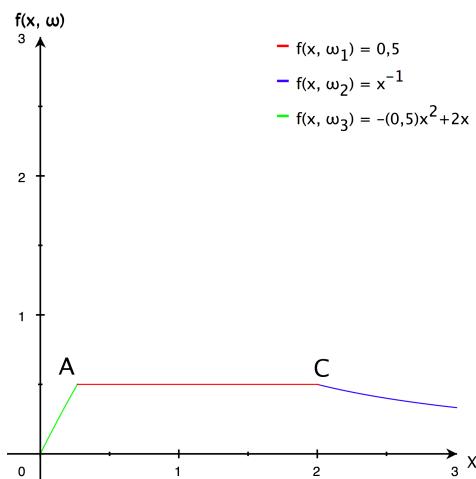


Figura 9.7: Applicazione del criterio del caso ottimo all'esempio continuo

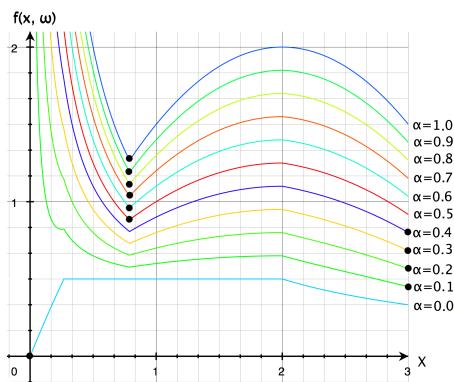


Figura 9.8: Applicazione del criterio di Hurwicz all'esempio continuo

α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$x_{\text{Hurwicz}}^*(\alpha)$	x_B	x_B	x_B	x_B	x_B	x_B	3	3	3	3	$\frac{1}{10^9}$

Tabella 9.28: Soluzioni ottime secondo il criterio di Hurwicz per ciascun valore campionario di α

Criterio di Laplace Il criterio consiste nel tracciare la curva media fra le tre date:

$$\phi_{\Omega}(x) = \frac{f(x, \omega^{(1)}) + f(x, \omega^{(2)}) + f(x, \omega^{(3)})}{3}$$

e calcolarne il minimo. La Figura 9.9 riporta l'andamento di tale curva in nero, evidenziandone il minimo, cioè $x_{\text{Laplace}}^* = 3$.

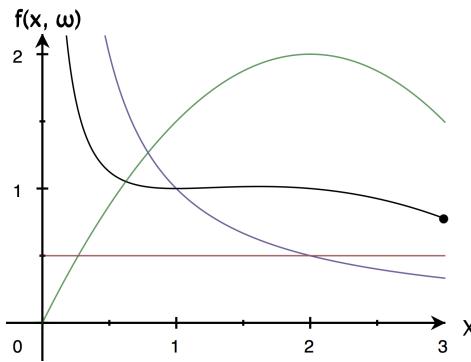


Figura 9.9: Applicazione del criterio di Laplace all'esempio continuo

Criterio del rammarico Per calcolare la funzione rammarico, occorre sottrarre a ciascuna funzione $f(x, \omega)$ il valore ottimo per lo scenario ω :

$$\rho(x, \omega) = f(x, \omega) - \min_{x \in X} f(x, \omega) \text{ per ogni } \omega \in \Omega$$

Le tre componenti della funzione rammarico sono riportate nella Tabella 9.29. La Figura 9.10 ne riporta i grafici.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
$f(x, \omega)$	$\frac{1}{2}$	$\frac{1}{x}$	$-\frac{1}{2}x^2 + 2x$
$f(x^*(\omega), \omega)$	$\frac{1}{2}$	$\frac{1}{3}$	10^{-18}
$\rho(x, \omega)$	0	$\frac{1}{x} - \frac{1}{3}$	$2x - \frac{1}{2}x^2$

Tabella 9.29: Valori della funzione rammarico per l'esempio continuo

Per concludere, bisogna individuare la curva del caso peggiore, cioè l'inviluppo superiore delle tre curve, ovvero i valori massimi per ogni $x \in X$. Nella figura, questa curva è ricalcata in nero. Il suo punto di ottimo è $x_{\text{Savage}}^* = x_D$.

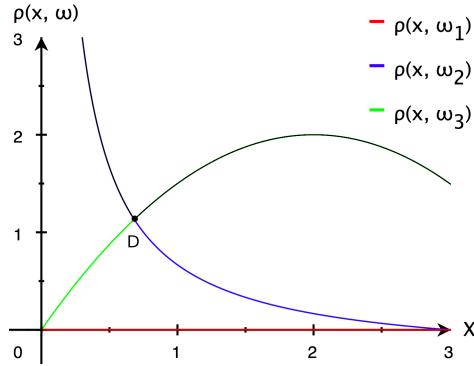


Figura 9.10: Andamento delle componenti della funzione rammarico per l'esempio continuo

Criterio delle eccedenze Per calcolare la funzione eccedenza, occorre sottrarre al valore pessimo per ciascuno scenario ω l'impatto $f(x, \omega)$:

$$\sigma(x, \omega) = \max_{x \in X} f(x, \omega) - f(x, \omega) \text{ per ogni } \omega \in \Omega$$

Le tre componenti della funzione eccedenza sono riportate nella Tabella 9.30. La Figura 9.11 ne riporta i grafici.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
$f(x^\dagger(\omega), \omega)$	$\frac{1}{2}$	10^9	2
$f(x, \omega)$	$\frac{1}{2}$	$\frac{1}{x}$	$-\frac{1}{2}x^2 + 2x$
$\sigma(x, \omega_i)$	0	$10^9 - \frac{1}{x}$	$\frac{1}{2}x^2 - 2x + 2$

Tabella 9.30: Valori della funzione eccedenza per l'esempio continuo

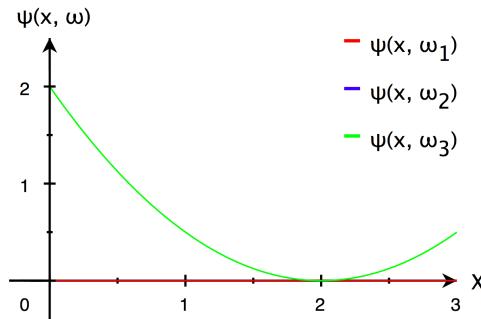


Figura 9.11: Andamento delle componenti della funzione eccedenza per l'esempio continuo

Per concludere, bisogna individuare la curva del caso peggiore, cioè l'inviluppo inferiore delle tre curve, ovvero i valori minimi per ogni $x \in X$. Questo inviluppo coincide con la componente $\sigma(x, \omega^{(1)}) = 0$. Tutte le soluzioni ammissibili $x \in X$ sono quindi indifferenti rispetto al criterio delle eccedenze.

9.8 Difetti formali delle logiche di scelta

Da una logica soddisfacente si richiederebbe che godesse delle seguenti proprietà:

1. *ordinamento debole*: dovrebbe indurre sempre un ordinamento debole sulle alternative;
2. *indipendenza dall'etichettatura*: l'ordinamento indotto non dovrebbe mutare se si permutano le alternative o gli scenari, ovvero si cambiano i loro nomi⁸;
3. *invarianza di scala*: l'ordinamento non dovrebbe mutare applicando all'impatto trasformazioni lineari $f' = \alpha f + \beta$ con $\alpha > 0$ e β qualsiasi, cioè misurando l'impatto con scale diverse (unità di misura e posizione dello zero);
4. *conservazione della dominanza forte*: se x domina x' , la logica dovrebbe essere far preferire x a x' almeno debolmente:

$$x \prec x' \Rightarrow \phi_{\Omega}(x) \leq \phi_{\Omega}(x')$$

5. *indipendenza dalle alternative irrilevanti*: l'ordinamento non dovrebbe mutare se si aggiungono o si togliano alternative (*rank reversal*);
6. *indipendenza dalla duplicazione di scenari*: l'ordinamento non dovrebbe mutare se si aggiungono scenari identici a quelli dati;
7. *indipendenza dalla variazione uniforme di uno scenario*: l'ordinamento delle alternative non dovrebbe mutare se gli impatti che corrispondono a uno stesso scenario migliorano o peggiorano uniformemente per tutte le alternative.

Purtroppo, nessuna delle logiche di scelta su elencate rispetta tutte le proprietà desiderabili. Peggio ancora, nessuna delle logiche di scelta possibili le rispetta. Quindi, non esiste nessuna motivazione razionale di fondo per preferire una logica a un'altra.

In particolare, le logiche su elencate godono delle prime quattro proprietà, ma non di tutte le altre. Per esempio, la prima proprietà deriva automaticamente dal fatto che si trasforma l'impatto $f(x, \omega)$ in una funzione $\phi_{\Omega}(x)$ e si ordinano le alternative in base ai valori numerici di tale funzione (necessariamente con un ordinamento debole). Consideriamo in dettaglio le ultime tre proprietà, facendo anche riferimento all'esempio con quattro alternative e quattro scenari discusso in precedenza.

9.8.1 Dipendenza da alternative irrilevanti

Le logiche del pessimismo, dell'ottimismo, di Hurwicz e di Laplace rispettano questa proprietà. Se si aggiunge una nuova alternativa, infatti, essa non fa che inserirsi nell'ordinamento delle vecchie in funzione del suo punteggio. Le vecchie soluzioni non cambiano ordine perché il loro punteggio rimane lo stesso: i valori del criterio $\phi_{\Omega}(\bar{x})$ per una specifica alternativa \bar{x} , infatti, dipendono solo dai valori dell'impatto $f(\bar{x}, \omega)$ per tale alternativa. La Tabella 9.31 riporta il calcolo di questi quattro criteri per il nuovo problema. È facile rendersi conto che i valori per le quattro alternative originali sono identici a quelli riportati nelle Tabelle 9.2, 9.3, 9.4 e 9.5, dato che vengono calcolati trattando ogni riga indipendentemente dalle altre.

⁸Esempi banali di criteri che violano questa condizione sono: "scegliere la prima alternativa" e "scegliere l'alternativa migliore nel primo scenario".

x	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{worst}}(x)$	$\phi_{\text{best}}(x)$	$\phi_{\text{Hurwicz}}(x)$	$\phi_{\text{Laplace}}(x)$
$x^{(1)}$	2	2	4	3	4	2	3.2	2.75
$x^{(2)}$	3	3	3	3	3	3	3.0	3.00
$x^{(3)}$	4	0	4	6	6	0	3.6	3.50
$x^{(4)}$	3	1	4	4	4	1	2.8	3.00
$x^{(5)}$	6	4	0	4	6	4	3.6	3.50

Tabella 9.31: Applicazione dei criteri del pessimismo, dell'ottimismo, di Hurwicz e di Laplace al problema di esempio dopo l'aggiunta di una quinta alternativa: l'ordinamento delle prime quattro non cambia

La logica del rammarico e quella delle eccedenze, invece, violano la proprietà, perché le funzioni rammarico $\rho(x, \omega)$ e $\sigma(x, \omega)$ per ciascuna x dipendono anche dalle alternative ottima $x^*(\omega)$ e pessima $x^\dagger(\omega)$ in ciascuno scenario ω . Siccome la nuova alternativa può risultare ottima o pessima in alcuni scenari e non in altri, il rammarico e l'eccedenza di altre alternative aumentano (non possono diminuire!) in tali scenari, mentre non cambiano negli altri scenari. Siccome rispetto al rammarico e all'eccedenza si considera lo scenario pessimo, i criteri ϕ_{Savage} e ϕ_{surplus} vedono cambiare alcuni valori, ma non altri, producendo quindi un ordinamento potenzialmente diverso.

x	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	2	2	4	3
$x^{(2)}$	3	3	3	3
$x^{(3)}$	4	0	4	6
$x^{(4)}$	3	1	4	4
$x^{(5)}$	6	4	0	4

Tabella 9.32: Applicazione del criterio di Savage al problema di esempio dopo l'aggiunta di una quinta alternativa: l'alternativa ottima per lo scenario $\omega^{(3)}$ cambia

La Tabella 9.32 riporta i costi del problema esteso con la soluzione $x^{(5)}$; in grassetto sono mostrate le alternative ottime per ogni scenario. Confrontandola con la Tabella 9.6, notiamo che $x^{(5)}$ è diventata la migliore alternativa per lo scenario $\omega^{(3)}$, sostituendo $x^{(2)}$. Di conseguenza, la funzione rammarico cambia (solo nella colonna $\omega^{(3)}$) e i suoi valori massimi riga per riga (cioè il caso pessimo per ciascuna alternativa) cambiano in maniera imprevedibile: alcuni rimangono identici, altri crescono dell'intera differenza fra il vecchio e il nuovo valore ottimo $f(x^*(\omega^{(3)}), \omega^{(3)})$, altri crescono di valori intermedi: la Tabella 9.33 illustra tutto questo. L'ordine risultante è $x^{(5)} \prec x^{(2)} \prec x^{(1)} \sim x^{(3)} \sim x^{(4)}$. Prima di aggiungere l'alternativa $x^{(5)}$, l'ordinamento delle altre quattro alternative era $x^{(4)} \prec x^{(1)} \prec x^{(2)} \sim x^{(3)}$, cioè diverso dall'attuale. Per esempio, $x^{(1)}$ era preferibile a $x^{(2)}$, mentre ora accade il contrario.

Analogamente, anche il criterio delle eccedenze viola questa proprietà. La Tabella 9.34 riporta i costi del problema esteso con la soluzione $x^{(5)}$. In grassetto sono evidenziate le alternative pessime per ogni scenario. Notiamo che $x^{(5)}$ è diventata l'alternativa pessima per lo scenario $\omega^{(1)}$, sostituendo $x^{(3)}$, e per lo scenario $\omega^{(2)}$, sostituendo $x^{(2)}$.

La Tabella 9.35 riporta i nuovi valori dell'eccedenza per ogni alternativa e scenario e, nell'ultima colonna, il valore nel caso pessimo per ciascuna alternativa e lo scenario corrispondente. La prima e la seconda colonna della matrice

$\rho(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{regret}}(x)$
$x^{(1)}$	0	2	4	0	4 ($\omega^{(3)}$)
$x^{(2)}$	1	3	3	0	3 ($\omega^{(2)}, \omega^{(3)}$)
$x^{(3)}$	2	0	4	3	4 ($\omega^{(3)}$)
$x^{(4)}$	1	1	4	1	4 ($\omega^{(3)}$)
$x^{(5)}$	4	1	0	1	1 ($\omega^{(1)}, \omega^{(2)}, \omega^{(4)}$)

Tabella 9.33: Valutazione del rammarico per il problema di esempio dopo l'aggiunta di una quinta alternativa

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	2	2	4	3
$x^{(2)}$	3	3	3	3
$x^{(3)}$	4	0	4	6
$x^{(4)}$	3	1	4	4
$x^{(5)}$	6	4	0	4

Tabella 9.34: Applicazione del criterio delle eccedenze al problema di esempio dopo l'aggiunta di una quinta alternativa

sono cambiate, crescendo della differenza fra il nuovo e il vecchio valore pessimo $f(x^\dagger(\omega_i), \omega_i)$ ($i = 1, 2$). Mentre prima le quattro alternative erano tutte indifferenti, ora l'alternativa $x^{(2)}$ risulta migliore delle altre.

$\sigma(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{surplus}}(x)$
$x^{(1)}$	4	2	0	3	0 ($\omega^{(3)}$)
$x^{(2)}$	3	1	1	3	1 ($\omega^{(2)}, \omega^{(3)}$)
$x^{(3)}$	2	4	0	0	0 ($\omega^{(3)}, \omega^{(4)}$)
$x^{(4)}$	3	3	0	2	0 ($\omega^{(3)}$)
$x^{(5)}$	0	0	4	2	0 ($\omega^{(1)}, \omega^{(2)}$)

Tabella 9.35: Valutazione dell'eccedenza per un problema di esempio con cinque alternative e quattro scenari

9.8.2 Dipendenza da duplicazioni di scenari

Quasi tutti i criteri analizzati risultano immuni agli sdoppiamenti di scenario, perché lo scenario ottimo o pessimo, anche se duplicati, rimangono identici, cioè i valori ottimo e pessimo dell'impatto per ogni alternativa non cambiano. Se però si usa la logica dell'equiprobabilità, duplicare uno scenario equivale a raddoppiare (quasi) la probabilità di tale scenario, e quindi il suo peso sulla funzione obiettivo complessiva. Questo modifica i valori del criterio $\phi_\Omega(x)$, e quindi può modificare l'ordinamento delle alternative.

La Tabella 9.36 riporta un'estensione dell'esempio nella quale lo scenario $\omega^{(2)}$ è stato duplicato aggiungendo uno scenario ω'_2 con impatti identici per ciascuna alternativa. L'ordine risultante del criterio di Laplace è $x^{(1)} \sim x^{(4)} \preceq x^{(3)} \preceq x^{(2)}$, mentre prima della duplicazione era $x^{(1)} \preceq x^{(2)} \sim x^{(4)} \preceq x^{(3)}$: l'alternativa $x^{(3)}$ è migliorata, perché il suo costo nello scenario $\omega^{(2)}$ è particolarmente basso, e quindi aumentare il peso di tale scenario la favorisce rispetto alle altre.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\omega'^{(2)}$	ϕ_{Laplace}
$x^{(1)}$	2	2	4	3	2	2.6
$x^{(2)}$	3	3	3	3	3	3.0
$x^{(3)}$	4	0	4	6	0	2.8
$x^{(4)}$	3	1	4	4	1	2.6

Tabella 9.36: Applicazione del criterio di Laplace a un problema di esempio (minimizzazione dei costi) con quattro alternative e cinque scenari

9.8.3 Dipendenza da variazioni uniformi in uno scenario

Le logiche del caso pessimo e ottimo (e quindi la logica di Hurwicz) sono suscettibili alla modifica uniforme di tutti gli impatti di uno scenario. Infatti, tale modifica può far sì che tale scenario diventi lo scenario pessimo o ottimo per alcune delle alternative, o al contrario che smetta di esserlo.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	ϕ_{worst}	ϕ_{best}	ϕ_{Hurwicz}	ϕ_{Laplace}
$x^{(1)}$	2	2	4	3	2	2	2.75	3.75
$x^{(2)}$	3	3	3	3	3	3	3.00	4.00
$x^{(3)}$	4	0	4	6	6	0	3.6	5.2
$x^{(4)}$	3	1	4	4	3	3	2.8	4.6

Tabella 9.37: Applicazione dei criterio del pessimismo, dell'ottimismo, di Hurwicz e di Laplace a un problema di esempio (minimizzazione dei costi) con quattro alternative e quattro scenari, in cui gli impatti relativi a uno scenario subiscono una modifica uniforme rispetto alle alternative

Nell'esempio cui stiamo facendo riferimento aumentiamo di un valore uniforme ($\delta f = 4$) tutti gli impatti dello scenario $\omega^{(2)}$. La Tabella 9.37 riporta i nuovi valori dell'impatto, nonché i valori dei criteri del pessimismo, dell'ottimismo, di Hurwicz (con $\alpha = 0.6$) e di Laplace. Come si vede, alcuni valori restano invariati, altri crescono di $\delta f = 4$, altri ancora di valori intermedi fra 0 e δf . Gli ordinamenti delle alternative cambiano come segue:

- secondo il criterio del pessimismo, da $x^{(2)} \preceq x^{(1)} \sim x^{(4)} \preceq x^{(3)}$ a $x^{(4)} \preceq x^{(1)} \sim x^{(3)} \preceq x^{(2)}$;
- secondo il criterio dell'ottimismo, da $x^{(3)} \preceq x^{(4)} \preceq x^{(1)} \preceq x^{(2)}$ a $x^{(1)} \preceq x^{(2)} \preceq x^{(3)} \sim x^{(4)}$;
- secondo il criterio di Hurwicz, da $x^{(4)} \preceq x^{(2)} \preceq x^{(1)} \preceq x^{(3)}$ a $x^{(1)} \preceq x^{(4)} \preceq x^{(3)} \preceq x^{(2)}$.

L'ordinamento secondo il criterio di Laplace, invece, non cambia, perché l'aggiunta di una costante δf agli impatti di uno scenario corrisponde all'aggiunta del valore uniforme $\delta f/|\Omega|$ per ogni $x \in X$ al criterio di Laplace $\phi_{\text{Laplace}}(x)$. In questo caso, il criterio cresce di $\delta f/4 = 1$ per ogni $x \in X$. Quindi, l'ordinamento delle alternative non cambia.

Anche il criterio del rammarico e delle eccedenze rispettano questa proprietà. Infatti sia la funzione di rammarico, sia la funzione di eccedenza sono definite come differenze fra impatti di alternative diverse (quella corrente e quella ottima, oppure quella pessima e quella corrente) in uno stesso scenario. Poiché a entrambi gli impatti si somma la stessa costante, i valori di rammarico ed eccedenza non cambiano.

Sia $\tilde{f}(x, \omega) = f(x, \bar{\omega}) + \delta$ la nuova funzione impatto. Allora, le nuove funzioni rammarico $\tilde{\rho}(x, \omega)$ ed eccedenza $\tilde{\sigma}(x, \omega)$ coincidono esattamente con le vecchie:

$$\begin{aligned}\tilde{\rho}(x, \omega) &= \tilde{f}(x, \omega) - \tilde{f}(x^*(\omega), \omega) = (f(x, \omega) + \delta) - (f(x^*(\omega), \omega) + \delta) \\ &= f(x, \omega) - f(x^*(\omega), \omega) = \rho(x, \omega) \\ \tilde{\sigma}(x, \omega) &= \tilde{f}(x^\dagger(\omega), \omega) - \tilde{f}(x, \omega) = (f(x^\dagger(\omega), \omega) + \delta) - (f(x(\omega), \omega) + \delta) \\ &= f(x^\dagger(\omega), \omega) - f(x(\omega), \omega) = \rho(x, \omega)\end{aligned}$$

9.9 La Programmazione Robusta

I concetti teorici presentati sinora nel caso finito si estendono quasi tutti banalmente ai casi in cui X e Ω sono infiniti (magari continui) oppure combinatorici, cioè finiti, ma di cardinalità assai elevata. Tuttavia, la soluzione pratica di tali problemi pone difficoltà molto serie: non è più possibile scorrere uno per uno gli scenari e le alternative applicando le definizioni. Spesso anche solo conoscere il valore di un criterio $\phi_\Omega(\cdot)$ in una prefissata alternativa \bar{x} richiede di risolvere un problema di ottimizzazione, che può non essere banale. Ad esempio, il criterio del caso pessimo richiede di calcolare $\min_{\omega \in \Omega} f(x, \omega)$. Il criterio del rammarico richiede addirittura tre ottimizzazioni successive, di cui due parametriche:

1. calcolare la miglior soluzione per ogni scenario, $\max_{x \in X} f(x, \omega)$, che è un'ottimizzazione parametrica in $\omega \in \Omega$;
2. calcolare lo scenario con il rammarico massimo per ogni alternativa, $\max_{\omega \in \Omega} \rho(x, \omega)$, che è un'ottimizzazione parametrica in $x \in X$;
3. calcolare l'alternativa con il criterio minimo: $\min_{x \in X} \phi_{\text{Savage}}(x)$.

Recentemente, però, lo studio della Programmazione in Condizioni di Ignoranza per problemi di Ottimizzazione Combinatoria ha visto un grosso fermento. Trattandosi di un'altra comunità scientifica rispetto a quella dell'economia classica, il nome assegnato al filone di studi è diverso: si parla infatti di *Programmazione Robusta*.

9.9.1 Criteri adottati in Ottimizzazione Combinatoria

Lo studio della robustezza in Ottimizzazione Combinatoria si fonda, come quello della programmazione in condizioni di ignoranza, sul sostituire l'impatto $f(x, \omega)$ con funzioni ausiliarie che eliminino la dipendenza dalle variabili esogene ω . A causa del diverso ambito applicativo, le funzioni ausiliarie usate non sono esattamente le stesse. Alcune vengono a mancare, perché non avrebbero molto senso nello studio di problemi con applicazioni tecnologiche e con moltissime soluzioni. Ad esempio, tutti i criteri ottimisti, che considerano lo scenario ottimo per ogni alternativa o l'eccedenza rispetto all'alternativa pessima per ogni scenario, vengono considerati poco utili, perché i sistemi tecnologici richiedono garanzie forti di buon funzionamento, e quindi non ammettono l'affidarsi alla fortuna. Di conseguenza, non si ritrovano in Programmazione Robusta il criterio del caso ottimo, quello di Hurwicz e quello delle eccedenze. Il criterio di Laplace richiederebbe una media aritmetica su un numero esponenziale o su un insieme infinito di scenari, e quindi è computazionalmente impraticabile. I criteri residui si conservano, ma assumono nomi diversi. Inoltre, sono stati proposti anche criteri del tutto nuovi. I più studiati sono:

1. la *robustezza assoluta*, che è il criterio del caso pessimo

$$\phi_{\text{RA}}(x) = \max_{\omega \in \Omega} f(x, \omega)$$

2. la *deviazione robusta*, che è il criterio del rammarico

$$\phi_{\text{DR}}(x) = \max_{\omega \in \Omega} \rho(x, \omega) = \max_{\omega \in \Omega} \left[f(x, \omega) - \min_{x \in X} f(x, \omega) \right]$$

3. la *robustezza relativa*, che valuta il rapporto fra rammarico e valore ottimo:

$$\phi_{\text{RR}}(x) = \max_{\omega \in \Omega} \frac{\rho(x, \omega)}{\min_{x \in X} f(x, \omega)} = \max_{\omega \in \Omega} \frac{f(x, \omega) - \min_{x \in X} f(x, \omega)}{\min_{x \in X} f(x, \omega)}$$

La robustezza assoluta esprime un atteggiamento molto conservativo: ottimizzarla significa difendersi dalle situazioni peggiori. La deviazione robusta e la robustezza relativa, invece, vedono l'incertezza anche come un'opportunità da sfruttare, con un approccio di tipo *benchmark*, che si confronta con il meglio che si poteva fare: la deviazione robusta misura quanto migliorerebbero in valore assoluto le prestazioni se si potesse eliminare l'incertezza, la robustezza relativa valuta tale miglioramento in rapporto al costo stesso, come se si trattasse di una perdita percentuale.

Esempio 67 Per semplicità, confrontiamo le tre definizioni di robustezza su un esempio finito di piccole dimensioni, anziché combinatorico: $X = \{x^{(1)}, x^{(2)}, x^{(3)}\}$ e $\Omega = \{\omega^{(1)}, \omega^{(2)}, \omega^{(3)}\}$. La seguente tabella riporta i costi per ogni alternativa e scenario, nonché quelli delle alternative ottime per ogni scenario.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
$x^{(1)}$	8	10	9
$x^{(2)}$	2	14	7
$x^{(3)}$	12	12	1
$f(x^*(\omega), \omega)$	2	10	1

Se ne ricava la robustezza assoluta $\phi_{\Omega}(x)$ (in grassetto è evidenziato il valore della soluzione ottima).

x	$\phi_{\text{RA}}(x)$
$x^{(1)}$	10
$x^{(2)}$	14
$x^{(3)}$	12

Di seguito viene riportata la tabella del rammarico e i valori della deviazione robusta (in grassetto è evidenziata la soluzione ottima).

$\rho(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\phi_{\text{DR}}(x)$
$x^{(1)}$	6	0	8	8
$x^{(2)}$	0	4	6	6
$x^{(3)}$	10	2	0	10

Per la robustezza relativa, il rammarico $\rho(x, \omega)$ viene rapportato al valore ottimo di ogni scenario. La tabella seguente riporta il rammarico “relativo” e i valori della robustezza relativa (la soluzione ottima è in grassetto).

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\phi_{RR}(x)$
$x^{(1)}$	$6/2 = 3$	$0/10 = 0.0$	$8/1 = 8$	8
$x^{(2)}$	$0/2 = 0$	$4/10 = 0.4$	$6/1 = 6$	6
$x^{(3)}$	$10/2 = 5$	$2/10 = 0.2$	$0/1 = 0$	5

Logiche diverse indicano soluzioni diverse: la robustezza assoluta indica $x^{(1)}$, la deviazione robusta $x^{(2)}$, la robustezza relativa $x^{(3)}$. Nel primo caso, importa il fatto di spendere il meno possibile (10); nel secondo caso, importa il fatto di spendere il meno possibile in più rispetto alla soluzione ottima (6 in più); nel terzo importa il fatto di spendere il meno possibile in più e in rapporto alla soluzione ottima (“solo” 5 volte di più, cioè il sestuplo).

9.9.2 Risultati teorici di complessità

Come si è detto, nel caso combinatorico non è praticamente possibile applicare le definizioni per valutare i criteri di scelta. Questo non vuol dire che non esistano algoritmi più efficienti per ottimizzare tali criteri: vuol dire invece che occorre studiare volta per volta la complessità del problema. Si verifica che, partendo dallo stesso problema deterministico, a seconda del criterio scelto per trattare l’incertezza (caso pessimo, ottimo, ecc...) e della descrizione dell’incertezza scelta (a scenari o a intervalli, si veda la Sezione 8.1) si possono ottenere problemi di complessità molto diversa tra loro.

In linea generale, si osserva che:

1. la controparte robusta (ovvero incerta) di un problema è più complessa del problema originale (spesso è \mathcal{NP} -completa anziché polinomiale);
2. ottimizzare la deviazione robusta o la robustezza relativa è più complicato che ottimizzare la robustezza assoluta, dato che richiede di conoscere la soluzione ottima per ogni scenario;
3. la descrizione dell’incertezza a scenari dà luogo a problemi più complessi di quella a intervalli, dato che:
 - non tutte le combinazioni di valori ammissibili per le singole variabili esogene corrispondono a scenari ammissibili;
 - ogni soluzione può avere un diverso scenario pessimo, mentre nella descrizione a intervalli, spesso tutte le soluzioni condividono lo stesso scenario pessimo⁹.

Nel seguito, discutiamo a mo’ di esempio il problema del cammino minimo robusto, distinguendo la complessità e gli algoritmi risolutivi per i due approcci nella descrizione dell’incertezza (scenari e intervalli) e per le tre funzioni obiettivo (robustezza assoluta, deviazione robusta e robustezza relativa).

Il problema del cammino minimo robusto: incertezza a scenari*

Consideriamo un grafo orientato $G = (N, A)$ con due nodi $s, t \in N$ e una funzione di costo $c : A \times \Omega \rightarrow \mathbb{N}$ definita sugli archi. Tale funzione di costo è incerta,

⁹Questo succede in particolare quando l’incertezza riguarda i coefficienti di una funzione di costo lineare: il caso pessimo, allora, è quello in cui tutti i coefficienti hanno il valore massimo.

*Questa è una sezione di approfondimento, che non fa parte del programma d’esame.

cioè dipende non solo dall'arco, ma anche dallo scenario che si verificherà. Supponiamo di descrivere l'incertezza con scenari esplicitamente elencati. La versione deterministica del problema ammette notoriamente diversi algoritmi polinomiali.

Nel seguito mostreremo che la versione robusta del problema, con tutte e tre le funzioni obiettivo, è \mathcal{NP} -completa. La dimostrazione avviene riducendo un opportuno problema \mathcal{NP} -completo al problema del cammino minimo robusto, cioè mostrando che un algoritmo che risolvesse quest'ultimo risolverebbe necessariamente anche il primo.

Definizione 36 *Dato un insieme N di oggetti e una funzione $c : S \rightarrow \mathbb{N}$ che associa un costo intero ad ogni oggetto, il Partition Problem richiede di determinare se sia possibile dividere N in due sottoinsiemi C e $S \setminus C$ che abbiano lo stesso costo totale:*

$$\exists C \subset S : \sum_{i \in C} c_i = \sum_{i \in S \setminus C} c_i ?$$

La Figura 9.12 riporta un insieme per cui non è possibile effettuare la partizione, mentre la Figura 9.13 riporta un esempio per cui la partizione è possibile.

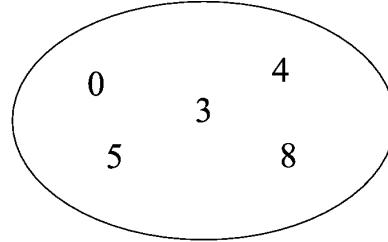


Figura 9.12: Insieme non partizionabile in due sottoinsiemi di ugual peso

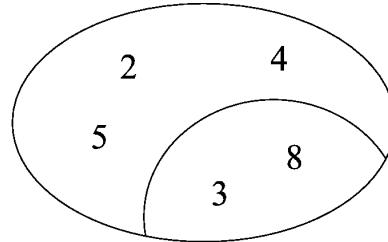


Figura 9.13: Insieme partizionabile in due sottoinsiemi di ugual peso

La Tabella 9.38 riporta tutti i possibili partizionamenti in due dell'insieme della Figura 9.13 e i costi dei due sottoinsiemi. La riga in grassetto mostra la partizione che soddisfa la proprietà richiesta, costituita dai due sottoinsiemi $\{3, 8\}$ e $\{2, 4, 5\}$.

Per ogni istanza del *Partition Problem*, si può costruire un'istanza del problema del cammino minimo robusto ad essa equivalente. Si consideri infatti un grafo $G = (N, A)$, in cui l'insieme dei nodi N contiene un nodo di partenza s , un nodo di arrivo v_{n+1} e $n = |S|$ coppie di nodi v_i e v'_i con $i = 1, \dots, n$, associate agli elementi dell'insieme S dato. Definiamo inoltre un insieme Ω con due scenari. Il nodo s è collegato al nodo v_1 da un arco il cui costo è 1 in entrambi gli scenari; ciascun nodo v_i ($i = 1, \dots, n$) è collegato al nodo v_{i+1} da un arco il cui costo è c_i nel primo scenario e 0 nel secondo e al nodo v'_i da un arco di costo 0 nel primo scenario e c_i

Partizione	Costi
$\emptyset - \{2, 3, 4, 5, 8\}$	0 – 22
$\{2\} - \{3, 4, 5, 8\}$	2 – 20
$\{3\} - \{2, 4, 5, 8\}$	3 – 19
$\{4\} - \{2, 3, 5, 8\}$	4 – 18
$\{5\} - \{2, 3, 4, 8\}$	5 – 17
$\{8\} - \{2, 3, 4, 5\}$	8 – 14
$\{2, 3\} - \{4, 5, 8\}$	5 – 17
$\{2, 4\} - \{3, 5, 8\}$	6 – 16
$\{2, 5\} - \{3, 4, 8\}$	7 – 15
$\{2, 8\} - \{3, 4, 5\}$	10 – 12
$\{3, 4\} - \{2, 5, 8\}$	7 – 15
$\{3, 5\} - \{2, 4, 8\}$	8 – 14
$\{3, 8\} - \{2, 4, 5\}$	11-11
$\{4, 5\} - \{2, 3, 8\}$	9 – 13
$\{4, 8\} - \{2, 3, 5\}$	12 – 10
$\{5, 8\} - \{2, 3, 4\}$	13 – 9

Tabella 9.38: Elenco delle partizioni in due sottoinsiemi dell’insieme di Figura 9.13

nel secondo. Ogni nodo v'_i è collegato al nodo v_{i+1} da un arco di costo nullo. La Figura 9.14 mostra il grafo equivalente al problema di partizione della Figura 9.13.

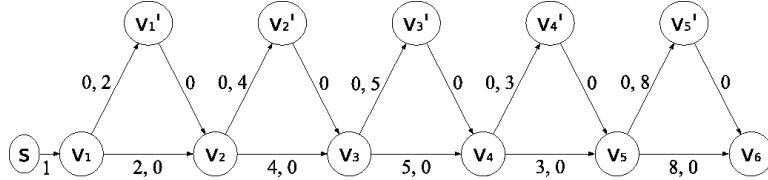


Figura 9.14: Grafo equivalente al problema partizione

Fissato uno scenario, il cammino minimo da s a v_{n+1} si ottiene in tempo $O(n)$, perché il grafo è aciclico e il numero di archi è proporzionale a quello dei nodi. È anche facile vedere che ogni cammino da s a v_{n+1} ha nel primo scenario un costo pari a 1 più la somma di alcuni c_i ($1 + \sum_{i \in C} c_i$), mentre nel secondo ha un costo pari a 1 più la somma degli altri c_i ($1 + \sum_{i \in N \setminus C} c_i$). Se si conoscesse lo scenario in anticipo, si potrebbe seguire sempre l’arco di costo nullo, e quindi ottenere un cammino di costo 1, ma il costo di tale cammino nell’altro scenario è pari a $1 + \sum_{i \in S} c_i$. Il numero di possibili soluzioni è 2^n ; infatti ci sono n nodi v_i in ciascuno dei quali si può scegliere tra due strade.

Si può osservare che, in realtà, il numero di alternative interessanti è 2^{n-1} , dato che metà delle soluzioni sono speculari all’altra metà: ogni soluzione x_j ($j = 1, \dots, 2^{n-1}$) ha una soluzione complementare $x_{j,c}$ in cui i valori per i due scenari sono invertiti e in cui in ogni nodo v_i ($i = 1, \dots, n$) viene selezionato l’arco uscente che non è stato visitato in x_j . Questo fatto si nota meglio nel *Partition Problem* originale, in cui non avevamo fatto distinzione tra il sottoinsieme C e il sottoinsieme $N \setminus C$ e, infatti, avevamo elencato $2^{(5-1)} = 16$ possibili partizioni.

La Tabella 9.39 riporta i $2^{(5-1)} = 16$ cammini interessanti, con i relativi costi nei due scenari. In particolare:

- $x^{(1)}$ è l’alternativa ottima per lo scenario $\omega^{(1)}$;

- $x^{(2)}$ è un'alternativa in cui il costo è lo stesso per i due scenari $\omega^{(1)}$ e $\omega^{(2)}$;
- altre quattordici alternative, $x^{(3)}, \dots, x_{16}$.

All'alternativa $x^{(1)}$ ottima per lo scenario $\omega^{(1)}$ corrisponde l'alternativa speculare $x'^{(1)} = (s - v_1 - v_2 - v_3 - v_4 - v_5 - v_6)$, che costa 23 nel primo scenario e 1 nel secondo, e quindi è ottima per lo scenario $\omega^{(2)}$. All'alternativa $x^{(2)}$ corrisponde l'alternativa speculare $x'^{(2)} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v_5 - v_6)$, che ha anch'essa costo 12 in entrambi gli scenari.

X	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{RA}(x)$
$x^{(1)} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	1	23	23
$x^{(2)} = (s - v_1 - v_2 - v_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	12	12	12
$x^{(3)} = (s - v_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	3	21	21
$x^{(4)} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v_5 - v'_5 - v_6)$	4	20	20
$x^{(5)} = (s - v_1 - v'_1 - v_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	5	19	19
$x_6 = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	6	18	18
$x_7 = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v_6)$	9	15	15
$x_8 = (s - v_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v_5 - v'_5 - v_6)$	6	18	18
$x_9 = (s - v_1 - v_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	7	17	17
$x_{10} = (s - v_1 - v_2 - v'_2 - v_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	8	16	16
$x_{11} = (s - v_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v_6)$	11	13	13
$x_{12} = (s - v_1 - v'_1 - v_2 - v_3 - v'_3 - v_4 - v_5 - v'_5 - v_6)$	8	16	16
$x_{13} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v_4 - v_5 - v'_5 - v_6)$	9	15	15
$x_{14} = (s - v_1 - v'_1 - v_2 - v_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	10	14	14
$x_{15} = (s - v_1 - v'_1 - v_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v_6)$	13	11	13
$x_{16} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v_4 - v'_4 - v_5 - v_6)$	14	10	14
\vdots	\vdots	\vdots	\vdots

Tabella 9.39: Elenco delle soluzioni interessanti del problema di cammino minimo di Figura 9.14

Vediamo in dettaglio che relazione c'è fra un cammino nel grafo e una partizione dell'insieme: in ogni cammino i pesi diversi da zero che vengono selezionati nello scenario $\omega^{(1)}$ sono quelli che appartengono al primo sottoinsieme della partizione associata; quelli che vengono selezionati con lo scenario $\omega^{(2)}$, invece, sono quelli che appartengono al secondo sottoinsieme. Ad esempio la partizione $\{5, 8\} - \{2, 3, 4\}$ corrisponde al cammino x_{16} . Notiamo che ogni alternativa contiene in entrambi gli scenari l'arco (s, v_1) , di peso 1, che non corrisponde ad alcun elemento nel problema di partizione. Tale arco serve solo per fare in modo che l'alternativa ottima abbia sempre costo maggiore di zero, così da evitare problemi nel denominatore della robustezza relativa. Questo spiega perché il costo di ogni cammino supera di un'unità il peso del sottoinsieme corrispondente nel problema della partizione.

Calcoliamo i tre criteri di robustezza per il problema di cammino minimo. Il valore $\phi_{RA}(x)$ della robustezza assoluta per ogni alternativa è riportato nell'ultima colonna della Tabella 9.39. L'alternativa migliore per la robustezza assoluta è $x^{(2)}$ (oppure l'alternativa complementare $x'^{(2)}$). Per costruzione, in questo specifico esempio la deviazione robusta e la robustezza relativa di ciascun cammino coincidono, poiché l'alternativa ottima costa 1 in ogni scenario, e quindi il denominatore presente nella robustezza relativa vale sempre 1. Possiamo, quindi, usare la funzione rammarico per minimizzare entrambi i criteri (vedi Tabella 9.40).

L'ultima riga della tabella riporta i valori della deviazione robusta e della robustezza relativa. L'alternativa che minimizza tale valore è $x^{(2)}$ (oppure la complemen-

f	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
$\omega^{(1)}$	0	11	2	3	4	5	8	5	6	7	10	7	8	9	12	13
$\omega^{(2)}$	22	11	20	19	18	17	14	17	16	15	12	15	14	13	10	9
ϕ_{DR}	22	11	20	19	18	17	14	17	16	15	12	15	14	13	12	13
ϕ_{RR}																

Tabella 9.40: Valori della deviazione robusta e della robustezza relativa per le soluzioni interessanti del problema di cammino minimo

tare $x'^{(2)}$). Quindi le tre funzioni obiettivo hanno lo stesso ottimo, che corrisponde a un cammino che distribuisce nel modo più equo i costi c_i . Un algoritmo polinomiale che ottimizzasse questo obiettivo sarebbe in grado di dire in tempo polinomiale se l'insieme dei c_i è divisibile in due parti uguali. Se nella soluzione ottima i due scenari non hanno ugual costo, allora l'insieme di partenza non è partizionabile in sottoinsiemi di ugual peso.

Il fatto che il *Partition Problem* sia \mathcal{NP} -completo, ma non fortemente \mathcal{NP} -completo lascia aperta la via a un algoritmo pseudopolinomiale. In effetti, esiste un algoritmo di programmazione dinamica che risolve il problema considerando nello spazio degli stati tutte le possibili durate dei cammini in ognuno dei due scenari. Se il numero di scenari è illimitato, però, il problema è fortemente \mathcal{NP} -completo, per riduzione dal *3-Partition Problem*.

Il problema del cammino minimo robusto: incertezza a intervalli¹⁰

Analizziamo ora il caso in cui l'incertezza sul costo di percorrenza di un arco (i, j) viene descritta a intervalli, cioè $c_{ij} \in [l_{ij}; \phi_{ij}]$. La Figura 9.15 mostra un grafo di esempio.

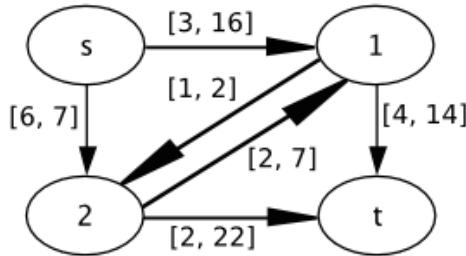


Figura 9.15: Problema del cammino minimo robusto, con incertezza descritta a intervalli

Ottimizzare la robustezza assoluta è un problema polinomiale, perché lo scenario pessimo è lo stesso per tutte le soluzioni ammissibili: si tratta dello scenario ω^\dagger in cui il costo di ciascun arco (i, j) è pari a ϕ_{ij} . Quindi basta applicare un classico algoritmo per il cammino minimo, assegnando a ogni arco il costo massimo ϕ_{ij} . La Tabella 9.41 riporta la robustezza assoluta di tutti i cammini semplici da s a t , evidenziando in grassetto l'alternativa prescelta.

¹⁰Questo è un argomento avanzato: non fa parte del programma d'esame. Vi si addentri chi è interessato a un approfondimento.

x	$\phi_{RA}(x)$
$s - 1 - t$	$16 + 14 = 30$
$s - 2 - t$	$7 + 22 = 29$
$s - 1 - 2 - t$	$16 + 2 + 22 = 40$
$s - 2 - 1 - t$	$7 + 7 + 14 = \mathbf{28}$

Tabella 9.41: Valori della robustezza assoluta per le soluzioni del problema di cammino minimo della Figura 9.15

Minimizzare la deviazione robusta è invece \mathcal{NP} -difficile, anche nel caso in cui il grafo sia orientato e aciclico¹¹. La deviazione robusta di un cammino x in uno scenario ω è il rammarico, cioè la differenza tra il costo $f(x, \omega)$ di x in ω e il costo $f(x^*(\omega), \omega)$ del cammino minimo nello scenario ω . È stato dimostrato in letteratura¹² che la deviazione robusta è massima nello scenario in cui i costi degli archi sono:

- $c_{ij} = \phi_{ij}$ se $(i, j) \in x$,
- $c_{ij} = l_{ij}$ se $(i, j) \notin x$.

Questo vuol dire che, nonostante l'insieme dei possibili scenari sia infinito e continuo, occorre valutare solo quelli in cui gli archi assumono valori estremi di costo, che sono in numero finito, benché esponenziale. Anzi, gli scenari potenzialmente interessanti sono solo quelli in cui gli archi di un cammino da s a t assumono il costo minimo e tutti gli altri il costo massimo.

Una possibile formulazione del problema è:

$$\min \sum_{(i,j) \in A} \phi_{ij} x_{ij} - d_t \quad (9.1)$$

$$\sum_{(i,j) \in A} x_{ij} - \sum_{(j,k) \in A} x_{jk} = \begin{cases} -1 & \text{per } j = s \\ 1 & \text{per } j = t \\ 0 & \text{altrimenti} \end{cases} \quad (9.2)$$

$$d_s = 0 \quad (9.3)$$

$$d_j \leq d_i + l_{ij} + (\phi_{ij} - l_{ij}) x_{ij} \quad (i, j) \in A \quad (9.4)$$

$$x_{ij} \in A \quad (9.5)$$

$$d_i \geq 0 \quad (9.6)$$

$$x_{ij} \in \{0, 1\} \quad (i, j) \in A \quad (9.7)$$

dove le variabili binarie x_{ij} identificano gli archi del cammino robusto, mentre d_t è il costo del cammino descritto dalle x_{ij} nello scenario pessimo. Ai vincoli classici imposti sulle variabili x_{ij} perché descrivano un cammino si aggiungono i nuovi vincoli, che impongono al costo effettivo di ogni arco (i, j) , cioè alla differenza $d_j - d_i$, di essere pari a l_{ij} se l'arco non fa parte del cammino descritto dalle x e pari a ϕ_{ij} se ne fa parte.

Effettuiamo il calcolo sull'esempio precedente. Poiché ci sono quattro cammini, dobbiamo valutare quattro scenari. Il $\sum_{(i,j) \in A} \phi_{ij} x_{ij}$ ci viene fornito dalla tabella della robustezza assoluta.

¹¹ Averbakh I, Lebedev V Interval data minmax regret network optimization problems. *Discrete Applied Mathematics* 138: 289–301, 2004

¹² Karaşan OE, Pinar MÇ, Yaman H The robust shortest path problem with interval data. *Operations Research Letters* 29: 31–40, 2001.

Per il cammino $(s - 1 - t)$ lo scenario che dobbiamo considerare è ω_{s-1-t} che viene mostrato in Figura 9.16.

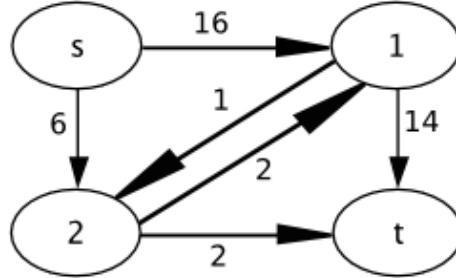


Figura 9.16: Scenario ω_{s-1-t}

Il cammino minimo per lo scenario ω_{s-1-t} è $(s - 2 - t)$ il cui costo è $6 + 2 = 8$. Quindi il valore della deviazione robusta per $s - 1 - t$ è $DR(s - 1 - t) = RA(s - 1 - t) - 8 = 30 - 8 = 22$. Per il cammino $(s - 2 - t)$ lo scenario che dobbiamo considerare è ω_{s-2-t} che viene mostrato in Figura 9.17.

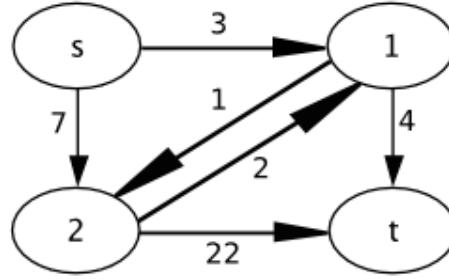


Figura 9.17: Scenario ω_{s-2-t}

Il cammino minimo per lo scenario ω_{s-2-t} è $(s - 1 - t)$ il cui costo è $3 + 4 = 7$. Quindi $DR(s - 2 - t) = RA(s - 2 - t) - 7 = 29 - 7 = 22$. Per il cammino $(s - 1 - 2 - t)$ lo scenario che dobbiamo considerare è $\omega_{s-1-2-t}$ che viene mostrato in Figura 9.18.

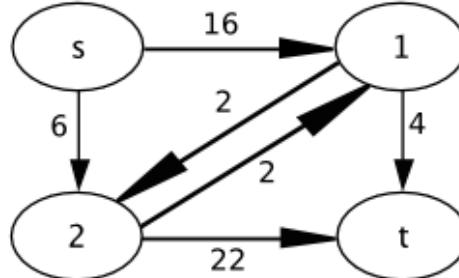


Figura 9.18: Scenario $\omega_{s-1-2-t}$

Il cammino minimo per lo scenario $\omega_{s-1-2-t}$ è $(s - 1 - t)$ il cui costo è $16 + 4 = 20$. Quindi $DR(s - 1 - 2 - t) = RA(s - 1 - 2 - t) - 20 = 40 - 20 = 20$. Per il cammino

$(s - 2 - 1 - t)$ lo scenario che dobbiamo considerare è $\omega_{s-2-1-t}$ che viene mostrato in Figura 9.19.

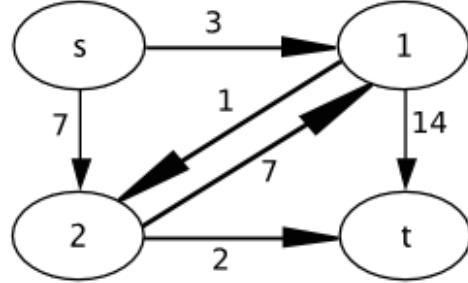


Figura 9.19: Scenario $\omega_{s-2-1-t}$

Il cammino minimo per lo scenario $\omega_{s-2-1-t}$ è $(s - 1 - 2 - t)$ il cui costo è $3 + 1 + 2 = 6$. Quindi $DR(s - 2 - 1 - t) = RA(s - 2 - 1 - t) - 6 = 28 - 6 = 22$. La Tabella 9.42 riassume i risultati: il cammino con deviazione robusta minima è $x^* = (s - 1 - 2 - t)$.

x	$\phi_{DR}(x)$
$(s - 1 - t)$	22
$(s - 2 - t)$	22
$(s - 1 - 2 - t)$	20
$(s - 2 - 1 - t)$	22

Tabella 9.42: Valutazione della deviazione robusta per il problema del cammino minimo con incertezza descritta a intervalli

Capitolo 10

Programmazione in condizioni di rischio

Rispetto alla programmazione in condizioni di ignoranza, il modello include oltre all'insieme Ω dei possibili scenari, una formalizzazione della loro probabilità. L'impatto $f(x, \omega)$ associato ad ogni soluzione è una *variabile aleatoria* che dipende dallo scenario ω , e si suppone che:

- nel caso discreto, sia nota una funzione che assegna a ciascuno scenario un valore di *probabilità* π_ω

$$\pi : \Omega \rightarrow [0; 1] \text{ con } \sum_{\omega \in \Omega} \pi_\omega = 1$$

- nel caso continuo, sia nota una funzione che assegna a ciascuno scenario un valore di *densità di probabilità* $\pi(\omega)$

$$\pi : \Omega \rightarrow \mathbb{R}^+ \text{ con } \int_{\Omega} \pi_{\omega} = 1$$

Per alcuni brevi richiami di teoria della probabilità si veda l'Appendice C.

Come nella programmazione in condizioni di ignoranza, gli approcci principali consistono nel ricondurre il problema all'ottimizzazione di una funzione ausiliaria $u(x)$, facendo sparire la dipendenza dallo scenario ω . Grazie alle ipotesi più forti, però, è possibile costruire criteri di decisione che godono delle proprietà che ci si attende da un approccio razionale.

10.1 Criterio del valore atteso

Le prime indagini su questi problemi, compiute nel Seicento da Pascal¹, si basano sulla rappresentazione approssimata dell'impatto di un'alternativa con il suo valore atteso:

$$u(x) = E[f(x, \omega)]$$

vale a dire

- nel caso discreto:

$$u(x) = E[f(x, \omega)] = \sum_{\omega \in \Omega} \pi_\omega f(x, \omega)$$

¹Blaise Pascal (1623-1662), matematico, fisico e filosofo francese.

- nel caso continuo:

$$u(x) = E[f(x, \omega)] = \int_{\Omega} \pi(\omega) f(x, \omega) d\omega$$

Esempio 68 Si consideri il seguente problema di decisione in condizioni di rischio: sono possibili quattro alternative e quattro scenari. La Tabella 10.1 riporta la matrice di valutazione con gli impatti per ogni alternativa e ogni scenario, ma in aggiunta riporta anche la probabilità che ciascuno scenario si realizzi, ovvero il vettore π .

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	1	3	4	6
$x^{(2)}$	2	2	2	4
$x^{(3)}$	3	2	1	9
$x^{(4)}$	6	6	1	3
$\pi^T(\omega)$	0.20	0.25	0.50	0.05

Tabella 10.1: Matrice di valutazione e vettore delle probabilità per un problema di programmazione in condizioni di rischio con quattro alternative e quattro scenari

Il criterio del valore atteso richiede di calcolare il valore atteso per ciascuna alternativa $x \in X$. Nel caso finito, questo corrisponde al prodotto della matrice di valutazione per il vettore delle probabilità:

$$u = U \cdot \pi \Rightarrow \begin{cases} u(x^{(1)}) = 1 \cdot 0.20 + 3 \cdot 0.25 + 4 \cdot 0.50 + 6 \cdot 0.05 = 3.25 \\ u(x^{(2)}) = 2 \cdot 0.20 + 2 \cdot 0.25 + 2 \cdot 0.50 + 4 \cdot 0.05 = 2.10 \\ u(x^{(3)}) = 3 \cdot 0.20 + 2 \cdot 0.25 + 1 \cdot 0.50 + 9 \cdot 0.05 = 2.05 \\ u(x^{(4)}) = 6 \cdot 0.20 + 6 \cdot 0.25 + 1 \cdot 0.50 + 3 \cdot 0.05 = 3.35 \end{cases}$$

Ne risulta l'ordinamento $x^{(3)} \prec x^{(2)} \prec x^{(1)} \prec x^{(4)}$, che porta a scegliere l'alternativa $x^{(3)}$.

10.2 Sensitività rispetto alle probabilità

Le probabilità dei singoli scenari sono spesso ricavate in modo campionario oppure stimate a partire da modelli, e quindi non sono note con assoluta precisione. È spesso conveniente, quindi, valutare la dipendenza della soluzione suggerita da un criterio dalle probabilità degli scenari, per capire se un eventuale errore può tradursi in una scelta sbagliata, e di quanto. Si tratta di individuare nello spazio delle probabilità le regioni nelle quali ciascuna alternativa risulta ottima.

Definizione 37 Per spazio delle probabilità si intende la regione

$$\mathcal{P}_{\Omega} = \left\{ \pi_{\omega} \in [0; 1]^r : \sum_{\omega \in \Omega} \pi_{\omega} = 1, \pi_{\omega} \geq 0, \forall \omega \in \Omega \right\}$$

dove $r = |\Omega|$.

Definizione 38 Definiamo supporto probabilistico di una soluzione $x \in X$ per un dato criterio di scelta il sottoinsieme Ω_x dello spazio delle probabilità \mathcal{P}_{Ω} nel quale x risulta ottima secondo il criterio stesso.

Se l'alternativa scelta in base alle probabilità stimate π risulta ottima in una regione ampia intorno al punto π , si può stare abbastanza tranquilli. In caso contrario, bisognerebbe prendere in considerazione per analisi ulteriori le soluzioni circostanti, oppure bisognerebbe migliorare la stima delle probabilità. Valutare l'ampiezza del supporto probabilistico di un'alternativa può essere particolarmente utile nelle decisioni in cui le probabilità sono state stimate in modo molto approssimativo, cioè in quelle che tendono ad avvicinarsi alla programmazione in condizioni di ignoranza. Molte decisioni finanziarie ricadono in questo ambito, dato che le probabilità in tali casi sono spesso definite soggettivamente.

Esempio 69 Riprendiamo l'Esempio 68: l'alternativa $x^{(2)}$, suggerita dal criterio del valore atteso, ha un valore molto simile a quello di $x^{(3)}$, tanto da far dubitare che imprecisioni nel valore degli impatti o delle probabilità possano aver determinato la scelta stessa. È quindi opportuno valutare la sensitività del risultato. Per semplicità, non compiamo lo studio nell'intero spazio delle probabilità, che avrebbe tre dimensioni indipendenti, ma analizziamo solo la sensitività rispetto a una delle probabilità. Le altre non possono rimanere fissate, perché per definizione la somma delle probabilità è unitaria, ma si possono tenere fissate le loro dimensioni relative, cioè la frazione che ciascuna occupa nel complemento della probabilità che si sta studiando. Questa semplificazione potrebbe anche essere dovuta a una minor fiducia che si nutre nel valore stimato per uno scenario rispetto agli altri.

Valutiamo la sensitività dell'esempio precedente rispetto alla probabilità dello scenario $\omega^{(3)}$. Ponendo $\pi(\omega^{(3)}) = \alpha$, le probabilità degli altri scenari si devono dividere il complemento $1 - \alpha$ in proporzione. Se indichiamo con $\bar{\omega}$ lo scenario la cui probabilità varia e con $\bar{\pi}_{\omega}$ le probabilità iniziali dei vari scenari, i nuovi valori di probabilità $\pi'_{\omega}(\alpha)$ con $\omega \neq \bar{\omega}$ si ottengono dai valori iniziali π_{ω} moltiplicandoli per un fattore moltiplicativo $k(\alpha)$ tale da rendere la loro somma uguale ad $1 - \alpha$:

$$\sum_{\omega \neq \bar{\omega}} k(\alpha) \bar{\pi}_{\omega} = (1 - \alpha) \Rightarrow k(\alpha) = \frac{1 - \alpha}{\sum_{\omega \neq \bar{\omega}} k(\alpha) \bar{\pi}_{\omega}} = \frac{1 - \alpha}{1 - \bar{\pi}_{\bar{\omega}}}$$

Quindi le nuove probabilità sono:

$$\pi'_{\omega}(\alpha) = \begin{cases} \frac{\alpha}{1 - \bar{\pi}_{\bar{\omega}}} & \text{per } \omega = \bar{\omega} \\ \frac{1 - \alpha}{1 - \bar{\pi}_{\bar{\omega}}} \pi'_{\omega} & \text{per } \omega \neq \bar{\omega} \end{cases}$$

Nell'esempio in questione:

$$\begin{cases} \pi(\omega^{(1)}) = 0.20 \frac{(1 - \alpha)}{1 - 0.50} = \frac{4}{10} (1 - \alpha) \\ \pi(\omega^{(2)}) = 0.25 \frac{(1 - \alpha)}{1 - 0.50} = \frac{5}{10} (1 - \alpha) \\ \pi(\omega^{(4)}) = 0.05 \frac{(1 - \alpha)}{1 - 0.50} = \frac{1}{10} (1 - \alpha) \end{cases}$$

Ora si può applicare il criterio del valore atteso in modo parametrico, e osservare le prestazioni delle quattro alternative al variare di α :

$$\begin{cases} u_{\alpha}(x^{(1)}) = 1 \cdot \frac{4}{10} (1 - \alpha) + 3 \cdot \frac{5}{10} (1 - \alpha) + 4 \cdot \alpha + 6 \cdot \frac{1}{10} (1 - \alpha) = \frac{25}{10} + \frac{15}{10} \alpha \\ u_{\alpha}(x^{(2)}) = 2 \cdot \frac{4}{10} (1 - \alpha) + 2 \cdot \frac{5}{10} (1 - \alpha) + 2 \cdot \alpha + 4 \cdot \frac{1}{10} (1 - \alpha) = \frac{22}{10} - \frac{2}{10} \alpha \\ u_{\alpha}(x^{(3)}) = 3 \cdot \frac{4}{10} (1 - \alpha) + 2 \cdot \frac{5}{10} (1 - \alpha) + 1 \cdot \alpha + 9 \cdot \frac{1}{10} (1 - \alpha) = \frac{31}{10} - \frac{21}{10} \alpha \\ u_{\alpha}(x^{(4)}) = 6 \cdot \frac{4}{10} (1 - \alpha) + 6 \cdot \frac{5}{10} (1 - \alpha) + 1 \cdot \alpha + 3 \cdot \frac{1}{10} (1 - \alpha) = \frac{57}{10} - \frac{47}{10} \alpha \end{cases}$$

In realtà, per tracciare l'andamento del criterio di scelta nelle quattro alternative basta osservare che si tratta di funzioni lineari di α , e che quindi è possibile disegnarlo una volta noto il suo valore in due punti. Per $\alpha = 1$, basta copiare il valore dell'impatto nello scenario considerato: $f(x, \bar{\omega}) = [4 \ 2 \ 1 \ 1]^T$. Per $\alpha = 0$, invece, bisogna calcolare la componente del criterio di scelta che deriva dagli altri scenari e dividerla per $1 - \bar{\pi}_{\bar{\omega}}$: $[1.25 \ 1.10 \ 1.55 \ 2.85]^T / 0.50$.

La Figura 10.1 illustra l'andamento del criterio nelle quattro alternative: per bassi valori di α , cioè quando lo scenario $\omega^{(3)}$ è improbabile, prevale l'alternativa $x^{(2)}$, mentre per alti valori di α , cioè quando lo scenario $\omega^{(3)}$ è probabile, prevale l'alternativa $x^{(3)}$. Il punto di intersezione fra i due diagrammi è individuato dall'equazione

$$u_{\alpha}(x^{(2)}) = u_{\alpha}(x^{(3)}) \Rightarrow \frac{22}{10} - \frac{2}{10}\alpha = \frac{31}{10} - \frac{21}{10}\alpha \Rightarrow \alpha = \frac{9}{19} \approx 0.474$$

Questo conferma che, se la stima di probabilità è incerta, la scelta di $x^{(3)}$ non è necessariamente la migliore. Le alternative $x^{(1)}$ e $x^{(4)}$, invece, sono sconfitte, rispettivamente, da $x^{(2)}$ e $x^{(3)}$ per ogni valore di probabilità. Questo potrebbe far pensare che siano dominate, ma non è così: le due soluzioni sono peggiori solo finché le probabilità degli scenari $\omega^{(1)}$, $\omega^{(2)}$ e $\omega^{(4)}$ rimangono nei rapporti reciproci fissati al principio. Non si tratta quindi di dominanza.

In realtà, è poi anche possibile che un'alternativa non dominata abbia supporto probabilistico vuoto, cioè che il criterio del valore atteso la sia consideri inferiore ad altre per qualsiasi valore di probabilità, esattamente come una soluzione paretiana può avere supporto vuoto, e la ragione è esattamente la stessa: le combinazioni lineari dei valori dell'impatto non sono sufficienti a determinare tutte le soluzioni non dominate.

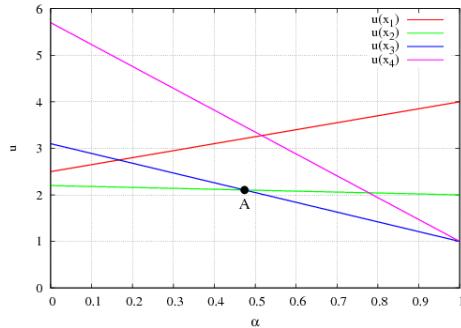


Figura 10.1: Studio di sensitività di un problema di minimizzazione di costi rispetto alla probabilità $\pi_{\omega^{(3)}} = \alpha$

10.3 Difetti formali del criterio del valore atteso

Sin dal Seicento, il criterio del valore atteso fu criticato per le conseguenze irrealistiche, e talvolta paradossali, a cui conduce. Passiamone in rassegna alcune.

10.3.1 Incoerenza fra valore atteso ed equivalente certo

Secondo il criterio del valore atteso, tutte le combinazioni di impatti e probabilità che producono lo stesso risultato sono vicendevolmente indifferenti. In pratica, si

osserva che non è così: se si propone a un decisore un impatto certo uguale al valore atteso degli impatti corrispondenti a diversi scenari, quasi sempre il decisore mostra una preferenza, talvolta a favore dell'impatto certo talaltra contro.

Esempio 70 *Si confrontino le seguenti 4 alternative:*

1. lanciare un dado e guadagnare con certezza 100 euro;
2. lanciare un dado e guadagnare 200 euro se escono 4, 5 o 6, nulla altrimenti;
3. lanciare un dado e guadagnare 600 euro se esce 6, nulla se escono gli altri numeri;
4. lanciare un dado e guadagnare 200 euro se escono 2, 3, 4, 5 o 6, perdere 400 euro se esce 1;

Secondo il criterio del valore atteso, le quattro soluzioni sono indifferenti: il valore atteso dell'impatto è infatti sempre di 100 euro. Eppure, quasi nessun decisore le considererebbe tali.

L'esempio precedente mostra che il valore atteso non modella correttamente la relazione di preferenza del decisore dal punto di vista descrittivo. Si potrebbe obiettare che questa è un'osservazione empirica sul comportamento umano, cioè riguarda i modelli descrittivi della preferenza umana (come i decisori agiscono in pratica) e non necessariamente i modelli prescrittivi (come i decisori dovrebbero agire per essere efficaci). Però anche dal punto di vista prescrittivo non sembra desiderabile che le quattro situazioni siano considerate indifferenti.

10.3.2 La scommessa di Pascal

Un famosissimo esempio di applicazione del criterio del valore atteso è la *scommessa di Pascal*, che egli intendeva come argomento per “dimostrare” l'esistenza di Dio, o meglio per sostenere la ragionevolezza della fede (non la sua razionalità logica). L'argomento, notissimo, afferma quanto segue. Ogni individuo deve scegliere se credere oppure no: possiamo modellare questa situazione con due alternative, $x^{(1)}$ e $x^{(2)}$. Inoltre, Dio esiste oppure non esiste, cioè vi sono due scenari. $\omega^{(1)}$ e $\omega^{(2)}$. La Tabella 10.2 fornisce la matrice di valutazione del problema. Se l'individuo crede e Dio esiste, egli guadagna il paradiso, che possiamo vedere come un beneficio A altissimo (al limite, infinito). I tre impatti associati al credere in un Dio che non esiste (e quindi perdere tempo, occasioni di godimento, ma anche vivere onestamente e serenamente, dice Pascal), al non credere in un Dio che esiste (e quindi essere, forse, punito) e non credere in un Dio che non esiste (probabilmente nessun impatto) vengono indicati rispettivamente con b , c e d , e sono tutti valori (positivi o negativi) molto inferiori ad A (in particolare, qualcuno potrebbe affermare che c sia un enorme valore negativo, corrispondente a una punizione infinita). Si noti che, trattandosi di benefici, il criterio del valore atteso deve essere massimizzato, anziché minimizzato.

$f(x, \omega)$	\exists Dio	\notin Dio
Credere	A	b
Non credere	c	d

Tabella 10.2: Matrice di valutazione degli impatti (benefici) per il problema della scommessa di Pascal: $A \gg |b|$, $A \gg c$ e $A \gg |d|$

Se indichiamo con α la probabilità (incognita) che Dio esista, il criterio del valore atteso genera le seguenti utilità per le due alternative:

$$\begin{cases} u(x^{(1)}) = A\alpha + b(1 - \alpha) \\ u(x^{(2)}) = c\alpha + d(1 - \alpha) \end{cases}$$

e la prima supera la seconda per $A\alpha + b(1 - \alpha) > c\alpha + d(1 - \alpha)$, cioè $(A - b - c + d)\alpha > d - b$. Se la ricompensa A è molto elevata (infinita), è ragionevole credere in Dio anche in base a una probabilità molto bassa; è sufficiente che sia:

$$\alpha > \alpha_{\min} = \frac{d - b}{A - b - c + d} = \frac{1}{1 + \frac{A - c}{d - b}}$$

Siccome $\lim_{A \rightarrow +\infty} \alpha_{\min} = 0$ per qualsiasi valore fissato di b, c e d , la possibilità di una ricompensa A infinita suggerisce la ragionevolezza della fede anche di fronte a una probabilità α estremamente piccola.

Obiezioni all'argomento furono avanzate immediatamente, oltre che sul piano della serietà etica (che però altri autori hanno difeso riformulando i termini della questione), anche sul piano della validità matematica. Si attaccò il fatto che le opzioni fossero due (vi sono molti modi di credere e non credere, e molte fedi in cui credere e non credere), che il concetto di probabilità in questa situazione non avesse senso (non si possono fare esperimenti casuali e vedere se al termine Dio esiste o no), ecc... Il punto che ci interessa qui, però, è l'uso del criterio del valore atteso, cioè l'ipotesi che il guadagno medio sia veramente ciò che sta a cuore a un essere umano.

10.3.3 Il paradosso di San Pietroburgo

Nel 1730, Bernouilli² propose il seguente gioco d'azzardo nel quale si vince sempre: si lancia una moneta finché non esce croce; se la prima croce esce dopo ω teste (con $\omega \in \mathbb{N}$), si vincono 2^ω euro. Quale somma c sarebbe ragionevole sborsare per poter partecipare al gioco?

Il problema prevede due alternative: non giocare o giocare. La prima offre un guadagno nullo. La seconda offre un guadagno pari a $-c + 2^\omega$, dove ω è il numero di teste che precedono la prima croce, e dipende ovviamente dallo scenario. Gli scenari possibili sono infiniti, uno per ogni possibile numero di teste consecutive ($\Omega = \mathbb{N}$). La probabilità dello scenario in cui escono ω teste seguite da una croce è $\pi_\omega = 1/2^{\omega+1}$. Quindi, il valore atteso del guadagno per la seconda alternativa è:

$$E[v] = -c + \sum_{\omega=0}^{+\infty} \frac{1}{2^{\omega+1}} 2^\omega = -c + \sum_{\omega=0}^{+\infty} \frac{1}{2} = +\infty$$

Quindi giocare è conveniente qualunque sia la cifra c , ovvero si sarebbe disposti a pagare qualsiasi cifra pur di partecipare al gioco. Non è chi non veda, però, che nessuno spenderebbe cifre molto elevate su questo gioco, perché sarebbe quasi certo di vincere solo cifre molto piccole. Entra in gioco un potentissimo fattore che viene detto *propensione al rischio* e che dipende dall'entità delle somme in gioco, dal reddito del giocatore, dal suo umore corrente, ecc...

Già Bernouilli propose una soluzione suggerendo che il valore atteso dovesse essere applicato non alla vincita, ma al logaritmo della vincita, in base al ragionamento

²Daniel Bernoulli (1700-1782), matematico e fisico svizzero, studioso di meccanica e fluidodinamica.

che ogni somma di denaro ha per un essere umano un valore che cala al crescere della sua ricchezza. Questo non risolve del tutto il paradosso, dato che si possono inventare giochi con vincite che crescono ancora più rapidamente. Solo un'utilità superiormente limitata risolve del tutto la questione.

10.4 Teoria dell'utilità stocastica

Negli anni '30 del Novecento, J. Von Neumann³ e O. Morgenstern⁴ introdussero la teoria dell'utilità stocastica, con l'obiettivo di superare i limiti del criterio del valore atteso attraverso un approccio assiomatico. L'idea di base è assumere che il decisore sia in grado di stabilire una relazione di preferenza Π non solo fra coppie di impatti deterministici, ma anche fra coppie di impatti incerti descritti come variabili aleatorie. Nel seguito, per semplicità ci concentreremo sul caso di insiemi di scenari finiti.

Definizione 39 Definiamo lotteria finita semplice una coppia di funzioni $\ell_{f,\pi} = (f(\omega), \pi(\omega))$ dove $f(\omega) : \Omega \rightarrow F$ è una variabile aleatoria su uno spazio campionario finito, mentre $\pi(\omega) : \Omega \rightarrow [0; 1]$ è la probabilità di ciascuno scenario in Ω . In particolare, si dice lotteria degenera una lotteria in cui uno scenario ha probabilità unitaria e gli altri hanno probabilità nulla, e lotteria binaria una lotteria in cui due scenari hanno probabilità di somma unitaria e gli altri hanno probabilità nulla.

L'insieme di tutte le lotterie semplici finite su F e Ω è quindi $F^{|\Omega|} \times \mathcal{P}_\Omega$, dato che contiene tutte le coppie di vettori in cui il primo ha $|\Omega|$ componenti in F e il secondo appartiene allo spazio delle probabilità su Ω . Siccome le lotterie degeneri corrispondono ai singoli impatti deterministici di F , indicheremo con ℓ_f qualsiasi lotteria degenera in cui l'unico impatto possibile sia f . Siccome le lotterie binarie hanno solo due impatti possibili, con probabilità complementari, e quindi la probabilità del primo determina automaticamente quella del secondo, anziché scrivere $\ell = ([f f']^T, [\alpha 1 - \alpha]^T)$, useremo la notazione $\ell = (f, \alpha, f')$.

Introdurremo poi il concetto di lotteria in più fasi, in cui ogni fase (tranne quella terminale) è una lotteria che mette in palio come premi dei biglietti per partecipare alla lotteria della fase immediatamente successiva; solo la fase terminale mette in palio vincite e perdite deterministiche.

Definizione 40 Definiamo lotteria composta una lotteria i cui impatti sono altre lotterie (eventualmente degeneri).

Definizione 41 Indichiamo con $L_{F,\Omega}$ l'insieme di tutte le lotterie, semplici e composte, definibili su F e Ω .

Dato un problema di decisione finito in condizioni di rischio e fissata una generica alternativa $\bar{x} \in X$, l'impatto $f(\bar{x}, \omega)$ e la probabilità π rispettano la definizione di lotteria finita. Quindi, ogni alternativa $x \in X$ corrisponde a una lotteria $\ell(x) \in L_{F,\Omega}$ e un metodo che consenta di confrontare lotterie consente anche di confrontare alternative.

³John Von Neumann, ovvero János Neumann (1903-1957), ebreo ungherese, trasferitosi negli Stati Uniti dalla Germania prima dell'avvento al potere dei nazisti, fu uno dei fondatori della teoria dei giochi e dell'informatica.

⁴Oskar Morgenstern (1902-1977), economista austriaco emigrato negli Stati Uniti dopo l'annessione dell'Austria alla Germania, uno dei fondatori della teoria dei giochi.

10.4.1 Assiomi fondativi dell'utilità stocastica

La teoria dell'utilità stocastica prima definisce le proprietà che una relazione di preferenza fra lotterie $\Pi \subset 2^{L \times L}$ dovrebbe rispettare per essere razionale. Quindi, la teoria dimostra che solo una ben determinata famiglia di relazioni di preferenza soddisfa tali proprietà, e che tali relazioni possono essere rappresentate attraverso delle funzioni conformi a valori reali. In questo modo, si riconduce la scelta di un'alternativa (lotteria) fra quelle ammissibili all'ottimizzazione della funzione conforme.

Definizione 42 *Data una relazione di preferenza fra lotterie $\Pi \subset 2^{L \times L}$, si dice che essa ammette una funzione conforme di utilità stocastica $u : L \rightarrow \mathbb{R}$ quando, per ogni coppia di lotterie ℓ e ℓ' , l'utilità della preferita supera l'utilità dell'altra:*

$$\ell \preceq \ell' \Leftrightarrow u(\ell) \geq u(\ell')$$

La teoria dimostra che tali funzioni coincidono tutte fra loro a meno di una riscalatura lineare, e quindi per ogni relazione di preferenza che rispetti le proprietà esiste una e una sola funzione di utilità stocastica normalizzata. La dimostrazione è costruttiva, cioè spiega anche in che modo costruire tale funzione.

Siccome ogni alternativa $x \in X$ di un problema di decisione in condizioni di rischio corrisponde a una lotteria $\ell(x)$, dotata di un'utilità stocastica $u(\ell(x))$, il problema si riconduce a trovare un'alternativa x° la cui corrispondente lotteria $\ell(x^\circ)$ abbia utilità stocastica massima rispetto a tutte le $x \in X$. Nel seguito, quindi, adottiamo il punto di vista degli economisti, che massimizzano benefici piuttosto che minimizzare costi.

Le proprietà richieste da una relazione di preferenza razionale fra lotterie, ovvero gli assiomi dell'utilità stocastica, sono i seguenti:

1. *ordinamento debole*: la relazione di preferenza fra lotterie Π è riflessiva, transitiva e completa;
2. *monotonia*: lotterie (semplici o composte) che assegnano probabilità più alte agli impatti o alle lotterie migliori sono preferite:

$$\alpha \geq \beta \Leftrightarrow (\ell, \alpha, \ell') \preceq (\ell, \beta, \ell') \text{ per ogni } \ell \preceq \ell'$$

3. *continuità*: date due lotterie ℓ e ℓ' , per qualsiasi impatto f intermedio fra ℓ e ℓ' esiste un opportuno valore di probabilità con il quale si possono comporre le due lotterie date in modo da ottenere una lotteria indifferente all'impatto:

$$\ell \preceq f \preceq \ell' \Rightarrow \exists \alpha \in [0; 1] : f \sim (\ell, \alpha, \ell')$$

È come dire che variando in modo continuo le probabilità delle lotterie, la preferenza varia in modo continuo, senza impatti che rimangono “scoperti”.

4. *indipendenza* (o *sostituzione*): la preferenza fra due lotterie non cambia se si aggiunge o si toglie la stessa lotteria con la stessa probabilità positiva:

$$f \preceq f' \Leftrightarrow (f, \alpha, f'') \preceq (f', \alpha, f'') \text{ per ogni } \alpha \in (0; 1]$$

5. *riduzione*: qualsiasi lotteria composta è indifferente alla lotteria semplice ottenuta elencando gli impatti distinti finali della lotteria composta e assegnando a ciascuno una probabilità data dalle regole classiche di composizione:

- la probabilità di ciascun singolo impatto finale è il prodotto delle probabilità incontrate lungo la sequenza di fasi della lotteria che conducono ad esso, dato che si tratta di probabilità condizionate;
- se un impatto finale compare più volte, generato da diversi rami della lotteria composta, le relative probabilità si sommano, dato che sono associate a eventi mutuamente incompatibili.

La Figura 10.2 illustra sulla sinistra la rappresentazione ad albero di una lotteria composta, ove ogni nodo corrisponde a una lotteria, ogni arco a uno scenario (con probabilità di somma unitaria per gli archi uscenti da un nodo) e ogni nodo foglia a un impatto finale. Nel caso in questione, la prima fase della lotteria ℓ ha due esiti possibili, che corrispondono alle lotterie ℓ_1 e ℓ_2 , rispettivamente con probabilità π_1 e π_2 (ovviamente $\pi_1 + \pi_2 = 1$). A sua volta, la lotteria ℓ_1 ha tre possibili esiti che corrispondono agli impatti f_1 (con probabilità π_{11}), f_2 (con probabilità π_{12}) e f_3 (con probabilità $\pi_{13} = 1 - \pi_{11} - \pi_{12}$), mentre la lotteria ℓ_2 ha due possibili esiti che corrispondono agli impatti f_1 (con probabilità π_{21}) e f_3 (con probabilità $\pi_{23} = 1 - \pi_{21}$). Sulla destra, la figura riporta la rappresentazione a stella della lotteria semplice equivalente, dove ogni arco lega la radice direttamente a un nodo foglia. I nodi foglia sono solo tre, perché diversi esiti della lotteria composta corrispondono allo stesso impatto finale. La probabilità su ogni arco è la somma, sui cammini che nell'albero originale legavano la radice a ciascuna foglia con un dato impatto, dei prodotti delle probabilità sugli archi di ciascun cammino. Per esempio, nell'albero originale due cammini legavano la radice all'impatto f_1 : il primo cammino attraversa un arco di probabilità π_1 e uno di probabilità π_{11} , mentre il secondo cammino attraversa un arco di probabilità π_2 e uno di probabilità π_{21} . Di conseguenza, nella stella finale la radice è collegata all'impatto f_1 da un arco con probabilità $\pi_1\pi_{11} + \pi_2\pi_{21}$.

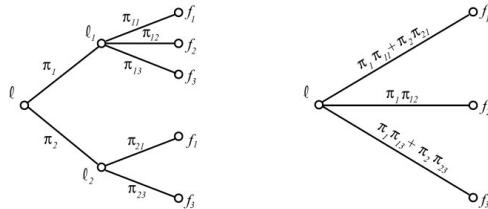


Figura 10.2: Rappresentazione ad albero di una lotteria composta, e sua sostituzione con una lotteria semplice equivalente

10.4.2 Il teorema dell'utilità stocastica di Von Neumann e Morgenstern

Teorema 22 *Dato un insieme di impatti F non tutti indifferenti tra loro, un insieme di scenari Ω e una relazione di preferenza Π fra lotterie su F e Ω che rispetti i cinque assiomi di Von Neumann e Morgenstern, esiste una e una sola funzione di utilità $u(\ell) : L_{F,\Omega} \rightarrow [0, 1]$ conforme a Π normalizzata in modo da assumere valore nullo nell'impatto pessimo e unitario in quello ottimo.*

Dimostrazione. Il primo assioma garantisce che la relazione di preferenza Π sia un ordine debole su tutte le lotterie di $L_{F,\Omega}$, e quindi anche sulle lotterie degeneri, cioè sugli impatti di F . Notiamo che questo assioma è assolutamente necessario ad avere una funzione di utilità conforme, dato che associare le lotterie a numeri reali

e la relazione di preferenza al confronto fra numeri reali significa automaticamente dotarla delle proprietà di un ordine debole. Non è però sufficiente: per costruire la funzione di utilità occorre un procedimento che applica via via anche gli altri assiomi.

L'esistenza di un ordine debole su F garantisce che ci sia almeno un impatto pessimo f^\dagger e almeno un impatto ottimo f^o in F . Se stiamo affrontando un problema di decisione, essi sono gli impatti estremi rispetto a tutte le variabili in gioco, sia x sia ω . Sia quindi $\ell^\dagger = (f^\dagger, 1)$ la lotteria degenera che propone con certezza l'impatto pessimo e $\ell^o = (f^o, 1)$ la lotteria degenera che propone con certezza l'impatto ottimo. Se gli impatti non sono tutti indifferenti fra loro, queste due lotterie sono certamente non indifferenti: $f^o \prec f^\dagger$, cioè $\ell^o \prec \ell^\dagger$. Ora possiamo assegnare valori convenzionali estremi all'utilità di queste due lotterie: utilità nulla alla lotteria degenera pessima ($u(\ell^\dagger) = 0$) e utilità unitaria alla lotteria degenera ottima ($u(\ell^o) = 1$).

Come secondo passo, si ricostruiscono i valori di utilità per tutte le lotterie degeneri, cioè per i singoli impatti, sfruttando l'assioma di continuità. Per ogni impatto $f \in F$ esiste certamente una probabilità $\alpha_f \in [0; 1]$ che produce una lotteria fra gli impatti estremi indifferente rispetto alla lotteria degenera corrispondente ad f :

$$\exists \alpha_f \in [0; 1] : (f^o, \alpha_f, f^\dagger) \sim f$$

Il valore α_f è unico per l'assioma di monotonia: infatti, se due lotterie $(f^o, \alpha_f, f^\dagger)$ e $(f^o, \beta_f, f^\dagger)$ fossero entrambe equivalenti a f , lo sarebbero anche fra loro, cioè ognuna sarebbe preferibile all'altra, e quindi avremmo sia $\alpha_f \leq \beta_f$ sia $\beta_f \leq \alpha_f$. In particolare, $\alpha_f = 1$ solo per lotterie degeneri ottime e $\alpha_f = 0$ solo per lotterie degeneri pessime.

Il terzo passo costruisce un valore di utilità per lotterie generiche. Per farlo, si sfruttano gli assiomi di sostituzione e di riduzione. Ogni possibile impatto finale $f \in F$ di una lotteria può essere visto come una lotteria degenera ℓ_f , che è equivalente a una lotteria binaria $(f^o, \alpha_f, f^\dagger)$ fra gli impatti estremi. Data una generica lotteria ℓ , si può sostituire ciascuno dei suoi impatti finali f con la lotteria degenera o con la lotteria binaria equivalente, ottenendo una lotteria composta a due fasi. La prima fase non fornisce più i singoli impatti, ma dei biglietti per partecipare alla seconda fase, nella quale sono possibili solo gli impatti estremi f^\dagger e f^o . Ora l'assioma di riduzione consente di combinare le due fasi in una lotteria semplice che ha gli stessi impatti finali (quelli estremi) e probabilità determinate dai teoremi della probabilità congiunta e della probabilità totale. In particolare, la probabilità dell'impatto ottimo è pari alla somma su ciascuno scenario $\omega \in \Omega$ del prodotto della probabilità π_ω dello scenario per la probabilità α_f dell'impatto ottimo nella lotteria che sostituisce f . A questo punto, la lotteria data ℓ corrisponde a una lotteria semplice fra gli impatti estremi:

$$\ell \sim \left(f^o, \sum_{\omega \in \Omega} \pi_\omega \alpha_{f(\omega)}, f^\dagger \right)$$

Definiamo l'utilità di tale lotteria ℓ come la probabilità dell'impatto ottimo f^o in essa. Siccome $\alpha_f = u(\ell_f)$, l'utilità stocastica della lotteria coincide con il valore atteso dell'utilità dell'impatto:

$$u(\ell) = \sum_{\omega \in \Omega} \pi_\omega u(f, 1) = E[u(f)]$$

■

Tornando dalle lotterie ai problemi di decisione in condizioni di rischio, ogni alternativa $x \in X$ corrisponde a una lotteria $\ell(x)$, costituita dall'impatto $f(x, \omega)$

di x e dalla probabilità π_ω . Il teorema di Von Neumann e Morgenstern consente di costruire un valore di utilità per $u(\ell(x))$, che possiamo direttamente vedere come una funzione di utilità $u(x)$ definita sulle alternative:

$$u(x) = \sum_{\omega \in \Omega} \pi_\omega u(f(x, \omega)) = E[u(f(x, \omega))] \quad x \in X$$

L'espressione dell'utilità stocastica è molto simile a quella del criterio del valore atteso ($u(x) = \sum_{\omega \in \Omega} \pi_\omega f(x, \omega) = E[f(x, \omega)]$), ma, anziché combinare gli impatti, combina le utilità delle corrispondenti lotterie degeneri. In altre parole, si tratta di una versione aggiornata e corretta del criterio del valore atteso, nel quale ciascun impatto non viene utilizzato direttamente, ma viene prima filtrato attraverso la funzione di utilità stocastica.

Esempio 71 Consideriamo un problema molto semplice, con quattro possibili impatti ($F = \{10, 20, 30, 50\}$), da intendere come benefici, e un dato insieme di scenari Ω . Costruiamo, seguendo i passi del teorema di Von Neumann e Morgenstern, una funzione di utilità stocastica che consenta di confrontare fra loro tutte le possibili lotterie finite $\ell = (f, \pi)$ definibili su F e Ω .

Siccome gli impatti sono benefici, l'impatto pessimo è il minore dei quattro e quello ottimo il maggiore: $f^\dagger = 10$ e $f^o = 50$. Le lotterie degeneri $\ell^\dagger = \ell_{f^\dagger}$ e $\ell^o = \ell_{f^o}$ hanno, rispettivamente, utilità $u(\ell^\dagger) = 0$ e $u(\ell^o) = 1$. L'assioma di continuità garantisce che ciascuno degli impatti intermedi $f^{(2)} = 20$ e $f^{(3)} = 30$ ammette una lotteria binaria equivalente fra gli impatti estremi, $(f^o, \alpha_{f^{(2)}}, f^\dagger)$ e $(f^o, \alpha_{f^{(3)}}, f^\dagger)$. Deve essere il decisore a indicare quale probabilità α garantisce l'indifferenza fra l'impatto intermedio e questa lotteria. Supponiamo che indichi $\alpha_{f^{(2)}} = 0.25$ e $\alpha_{f^{(3)}} = 0.60$, cioè che il decisore consideri indifferente guadagnare $f^{(2)} = 20$ con certezza oppure avere una probabilità pari a 0.25 di guadagnare $f^o = 50$ e una probabilità complementare pari a 0.75 di guadagnare $f^\dagger = 10$, e analogamente che consideri indifferente guadagnare $f^{(3)} = 30$ con certezza oppure avere una probabilità pari a 0.60 di guadagnare $f^o = 50$ e una probabilità complementare pari a 0.40 di guadagnare $f^\dagger = 10$. Quindi, $u(f^{(2)}, 1) = 0.25$ e $u(f^{(3)}, 1) = 0.60$.

A questo punto, l'utilità di qualsiasi lotteria fra i quattro impatti risulta automaticamente determinata. Per esempio, supponiamo che gli scenari $\omega^{(1)}$, $\omega^{(2)}$ e $\omega^{(3)}$ abbiano probabilità $\pi_{\omega^{(1)}} = 0.25$, $\pi_{\omega^{(2)}} = 0.50$ e $\pi_{\omega^{(3)}} = 0.25$ e paragoniamo le alternative x e x' , che producono i seguenti impatti: $f(x, \omega^{(1)}) = 10$, $f(x, \omega^{(2)}) = 20$, $f(x, \omega^{(3)}) = 50$ e $f(x', \omega^{(1)}) = 50$, $f(x', \omega^{(2)}) = 10$, $f(x', \omega^{(3)}) = 30$. Entrambe le alternative sono lotterie. Le loro utilità stocastiche possono essere calcolate come il valore atteso dell'utilità, cioè come le combinazioni convesse delle utilità associate ai singoli impatti con le probabilità degli scenari corrispondenti:

$$u(x) = \sum_{\omega \in \Omega} \pi_\omega f(x, \omega) = 0.25 \cdot 0 + 0.50 \cdot 0.25 + 0.25 \cdot 1 = 0.375$$

$$u(x') = \sum_{\omega \in \Omega} \pi_\omega f(x', \omega) = 0.25 \cdot 1 + 0.50 \cdot 0 + 0.25 \cdot 0.60 = 0.40$$

Poiché $u(x') > u(x)$, la seconda alternativa è preferibile alla prima: $x' \prec x$.

C'è in questo procedimento una forte analogia con la costruzione della funzione di utilità a molti attributi. Del resto, abbiamo già osservato una somiglianza nella definizione di dominanza: è come se gli scenari fossero trattati come attributi. D'altra parte, ci sono anche forti differenze. Per esempio, una relazione di preferenza fra impatti deterministici ammette infinite funzioni valore conformi (basta applicare una qualsiasi funzione monotona crescente). È solo nel caso additivo, sotto forti

ipotesi aggiuntive e con molta fatica modellistica, che si riesce a costruire una combinazione lineare delle singole utilità con opportuni pesi. Nel caso delle relazioni di preferenza fra lotterie, i cinque assiomi garantiscono che esiste una sola funzione di utilità conforme. Anche rilassando la condizione di normalizzazione, si ottengono infinite funzioni conformi, ma tutte legate da un semplice cambiamento di scala $u'(\ell) = au(\ell) + b$ con $a > 0$. Infatti, il teorema di Von Neumann e Morgenstern si può dimostrare sostituendo semplicemente i valori convenzionali 0 e 1 per gli impatti estremi con altri due valori reali. Inoltre, i coefficienti della combinazione lineare sono delle probabilità, spesso stimabili in base alle frequenze osservate degli scenari, e non dei tassi marginali di sostituzione, da stimare con interviste al decisore. Il punto è che in questo caso si stanno combinando quantità omogenee (l'impatto è un solo indicatore, misurato in situazioni diverse) e non eterogenee (impatto vettoriale composto da indicatori diversi).

10.5 Avversione e propensione al rischio

Ogni decisore ha una propria funzione di utilità stocastica $u : L_{F,\Omega} \rightarrow [0, 1]$, che riflette strettamente il suo atteggiamento verso l'incertezza. L'andamento di tale funzione indica anche quanto le indicazioni dell'utilità stocastica differiscono da quelle fornite dal classico criterio del valore atteso. Per descrivere l'andamento di $u(\ell)$ in modo semplice, ci limitiamo a considerare lotterie degeneri ℓ_f , con $f \in F$, cioè consideriamo la restrizione della funzione u dall'insieme delle lotterie $L_{F,\Omega}$ a quello degli impatti F . Questo semplifica il compito perché permette di tracciare un diagramma con un solo argomento f , mentre, anche solo considerando le lotterie semplici, si avrebbero $2r - 1$ argomenti: gli r impatti $f(\omega)$ e le r probabilità π_ω legate dal vincolo di somma unitaria. D'altra parte, come dimostra il teorema, i valori di u in F determinano univocamente quelli in tutto $L_{F,\Omega}$.

Definizione 43 Si definisce profilo di rischio l'andamento dell'utilità stocastica sulle lotterie degeneri ℓ_f al variare dell'impatto $f \in F$.

Se supponiamo, come abbiamo fatto sinora, che l'impatto f esprima un beneficio, il profilo di rischio è una funzione di f monotona strettamente crescente, che parte da $u(\ell_{f^\dagger}) = 0$ e arriva a $u(\ell_{f^\circ}) = 1$. Consideriamo due impatti f' e f'' , con $f'' > f'$ (dunque, $f'' \prec f'$). Qualsiasi impatto intermedio $f \in [f', f'']$ si può ottenere con un'opportuna combinazione convessa dei due impatti estremi: $f_\alpha = \alpha f'' + (1 - \alpha) f'$. Tale impatto è intermedio anche rispetto alla preferenza: $f'' \preceq f_\alpha \preceq f'$.

Ora consideriamo i seguenti tre casi:

1. il profilo di rischio $u(f)$ è funzione concava di f in $[f', f'']$:

$$u(\alpha f'' + (1 - \alpha) f', 1) \geq \alpha u(\ell_{f''}) + (1 - \alpha) u(\ell_{f'}) \text{ per ogni } \alpha \in [0; 1]$$

2. il profilo di rischio $u(f)$ è funzione lineare di f in $[f', f'']$

$$u(\alpha f'' + (1 - \alpha) f', 1) = \alpha u(\ell_{f''}) + (1 - \alpha) u(\ell_{f'}) \text{ per ogni } \alpha \in [0; 1]$$

3. il profilo di rischio $u(f)$ è funzione convessa di f in $[f', f'']$

$$u(\alpha f'' + (1 - \alpha) f', 1) \leq \alpha u(\ell_{f''}) + (1 - \alpha) u(\ell_{f'}) \text{ per ogni } \alpha \in [0; 1]$$

La Figura 10.3 descrive i tre casi. In ciascuno, gli estremi del profilo di rischio sono marcati dai punti $P_1 = (f', u(\ell_{f'}))$ e $P_2 = (f'', u(\ell_{f''}))$, mentre il punto

$P = (f, u(\ell_f))$ intermedio indica l'impatto f e la sua utilità stocastica. Questi tre casi non esauriscono tutte le possibilità, ma sono situazioni esemplari. In ciascuno dei tre diagrammi compare anche il punto Q che ha ascissa f_α e ordinata $\alpha u(f'', 1) + (1 - \alpha) u(f', 1)$ e giace sul segmento $\overline{P_1 P_2}$. Indaghiamo ora il significato di questo punto.

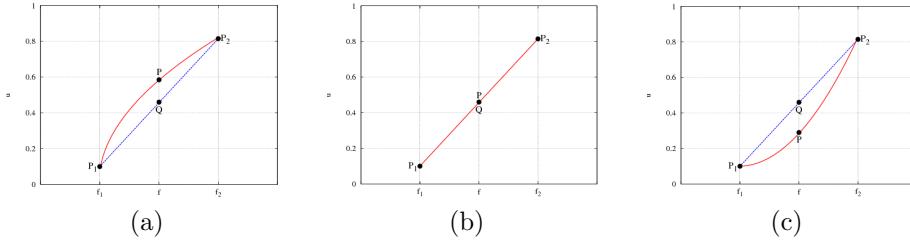


Figura 10.3: Profili di rischio di tre decisori: avverso al rischio (a), neutrale al rischio (b) e propenso al rischio (c)

Invece di un impatto deterministico intermedio fra f' e f'' , consideriamo una lotteria ℓ_α fra questi due impatti, che attribuisca a f'' una probabilità pari ad α . L'utilità stocastica di Von Neumann e Morgenstern della lotteria ℓ_α è:

$$u(\ell_\alpha) = \alpha u(f') + (1 - \alpha) u(f'')$$

che è l'ordinata del punto Q , mentre l'utilità stocastica di f_α , cioè della lotteria degenere ℓ_{f_α} , è $u(\ell_f)$, che è l'ordinata del punto P . Dunque, in generale $\ell_{f_{alpha}} \not\sim \ell_\alpha$.

Supponiamo invece di applicare il classico criterio del valore atteso per confrontare le due alternative ℓ_{f_α} e ℓ_α utilità di dell'impatto f . Siccome ℓ_{f_α} è una lotteria degenere, il valore del criterio è banalmente $\phi_{EV}(\ell_{f_\alpha}) = 1 \cdot f_\alpha = f_\alpha$. D'altra parte, il criterio del valore atteso per la lotteria $\ell_\alpha = (f'', \alpha, f')$ è $\phi_{EV}(\ell_\alpha) = \alpha f'' + (1 - \alpha) f' = f_\alpha$. Dunque, secondo il criterio del valore atteso $f \sim \ell_\alpha$.

A questo punto, i tre andamenti acquistano un significato intuitivo. Costruiti con lo stesso coefficiente α un impatto certo intermedio e una lotteria fra gli stessi impatti estremi, il criterio del valore atteso li considera indifferenti, mentre:

1. se il profilo di rischio $u(f)$ è funzione concava, il decisore preferisce l'impatto certo, dunque è *avverso al rischio*;
2. se il profilo di rischio $u(f)$ è funzione lineare, il decisore è indifferente, dunque è *neutrale al rischio*;
3. se il profilo di rischio $u(f)$ è funzione convessa, il decisore preferisce la lotteria, dunque è *propenso al rischio*.

In generale, un profilo di rischio abbastanza regolare sarà scomponibile in una sequenza di intervalli, in ognuno dei quali assume uno dei tre andamenti tipici. Ad esempio, un decisore può considerare gratificante giocare piccole somme anche se il valore atteso della vincita è negativo, cioè peggiore che non giocare, ma lo stesso decisore può rifiutare di giocare grandi somme persino quando il valore atteso della vincita è positivo. Viceversa, un altro decisore potrebbe amare il rischio solo per grandi somme ed evitarlo per le piccole.

10.5.1 Equivalente certo e premio di rischio

Una descrizione equivalente del profilo di rischio di un decisore è data dalla funzione inversa dell'utilità, cioè dalla funzione che ricostruisce un impatto $f(u)$ per ciascun

valore dell'utilità stocastica u . Questa funzione inversa esiste sicuramente, dato che $u(f)$ è strettamente crescente, e si ottiene scambiando ascisse e ordinate nel diagramma del profilo di rischio.

Definizione 44 *Data una lotteria ℓ , si dice equivalente certo l'impatto f_ℓ che le equivale, e premio di rischio la differenza fra il valore atteso della lotteria e il suo equivalente certo.*

Il premio di rischio misura l'utilità aggiuntiva che il decisore richiede per accettare una lotteria al posto del suo valore atteso. Quindi:

- per i decisori avversi al rischio, l'equivalente certo di una lotteria è inferiore al suo valore atteso, e quindi il premio di rischio è positivo;
- per i decisori neutrali al rischio, l'equivalente certo di una lotteria è identico al suo valore atteso, e quindi il premio di rischio è nullo;
- per i decisori propensi al rischio, l'equivalente certo di una lotteria è superiore al suo valore atteso, e quindi il premio di rischio è negativo.

Esempio 72 *Supponiamo che un problema di decisione abbia come regione degli impatti l'intervallo $[0; 1000]$ e che il profilo di rischio del decisore sia $u(f, 1) = \sqrt{f}/1000$. Se si dovesse scegliere fra una vincita sicura da $\bar{f} = 250$ euro e una lotteria $\ell = (f, \pi)$ che fornisce 810 euro con probabilità pari a 0.1, 360 euro con probabilità 0.5 e 160 euro con probabilità 0.4, quale alternativa si dovrebbe scegliere? Qual è l'equivalente certo della lotteria?*

L'utilità della vincita sicura è $u(\bar{f}) = \sqrt{\frac{250}{1000}} = 0.5$; l'utilità della lotteria è:

$$\begin{aligned} u(\ell) &= \sum_{\omega \in \Omega} \pi_\omega u(f(\omega)) = \\ &= 0.1 \cdot u(810) + 0.5 \cdot u(360) + 0.4 \cdot u(160) = \\ &= 0.1 \cdot 0.9 + 0.5 \cdot 0.6 + 0.4 \cdot 0.4 = 0.55 \end{aligned}$$

e quindi la lotteria è preferibile alla vincita sicura.

L'equivalente certo della lotteria è l'impatto f_ℓ che avrebbe utilità identica:

$$u(f_\ell) = \sqrt{\frac{f_\ell}{1000}} = 0.55 \Rightarrow f_\ell = 302.5$$

Per calcolare il premio di rischio, bisogna valutare il valore atteso della vincita della lotteria

$$E[f] = \sum_{\omega \in \Omega} \pi_\omega f(\omega) = 0.1 \cdot 810 + 0.5 \cdot 360 + 0.4 \cdot 160 = 81 + 180 + 64 = 325$$

Quindi, il premio di rischio è $E[f] - f_\ell = 325 - 302.5 = 23.5$, positivo. Infatti, il decisore ha un profilo di rischio concavo, e quindi è avverso al rischio: rispetto alla lotteria, preferisce una vincita certa più bassa del valore atteso della lotteria.

10.6 Critiche alla teoria dell'utilità stocastica*

Una critica abbastanza immediata alla teoria dell'utilità stocastica è che essa, come la teoria dell'utilità a molti attributi, richiede un decisore estremamente attento,

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

preciso e capace di soppesare le alternative propostegli, che non sono semplici impatti, ma lotterie ipotetiche. Criteri più semplici, come il bistrattato criterio del valore atteso, sono direttamente applicabili senza dover aggiungere ulteriori informazioni opinabili.

Un filone di critiche più fondamentale si indirizza invece contro gli stessi assiomi, e in particolare l'assioma di indipendenza. Questo richiede che la preferenza fra due lotterie non cambi se si aumenta o riduce in ugual misura la probabilità di impatti identici nelle due lotterie. Sembra un'ipotesi ragionevole, ma diversi esperimenti psicologici mostrano che non accade in pratica.

10.6.1 Paradosso di Allais

Questo paradosso, formulato da M. Allais⁵ nel 1953, propone la scelta tra:

1. una lotteria con premi $f = [2\ 500 \ 2\ 400 \ 0]^T$ e probabilità $\pi = [0.33 \ 0.66 \ 0.01]^T$
2. una lotteria degenere con premio certo pari a $f_2 = 2\ 400$.

I decisori più avversi al rischio preferiscono la seconda lotteria, cioè il premio certo.

Supponiamo ora di modificare entrambe le lotterie allo stesso modo, sostituendo in un sottoinsieme di scenari di probabilità 0.66 l'impatto f_2 con l'impatto f_3 . Questo equivale a eliminare tale sottoinsieme di scenari e aggiungere un altro sottoinsieme di uguale probabilità con un impatto diverso. Secondo l'assioma di sostituzione, la preferenza fra le due lotterie non dovrebbe cambiare. Nella prima lotteria, l'impatto f_2 viene sostituito da f_3 in tutti gli scenari dove compariva, mentre nella seconda viene sostituito in una parte. Il risultato sono le due lotterie seguenti:

1. una lotteria con premi $f = [2\ 500 \ 0]^T$ e probabilità $\pi = [0.33 \ 0.67]^T$
2. una lotteria con premi $f = [2\ 400 \ 0]^T$ e probabilità $\pi = [0.34 \ 0.66]^T$

Sperimentalmente, gli stessi decisori che preferivano la seconda lotteria ora scelgono in maggioranza la prima: la preferenza si è invertita.

Il fatto è che nella prima scelta l'impatto nullo, benché molto improbabile, impressiona i decisori e li spinge a cercare una certezza, mentre nel secondo test la probabilità di non vincere nulla è talmente alta da indurre i decisori a rischiare per guadagnare di più nel caso fortunato. In altre parole, 0.01 e 0 sembrano probabilità molto diverse, mentre 0.33 e 0.34 sembrano molto simili.

Il paradosso di Allais e altre considerazioni simili portarono D. Kahneman⁶ e A. Tversky⁷ a proporre la *teoria del prospetto*, nella quale

1. l'utilità degli impatti ha un andamento a forma di *S* asimmetrica, per cui il decisore soffre le perdite più di quanto goda dei guadagni di uguale entità;
2. la funzione valore combina linearmente i valori delle utilità non con le probabilità, ma con funzioni non lineari che sovrastimano le probabilità piccole e sottostimano quelle grandi.

Queste complicazioni consentono di descrivere meglio il comportamento dei decisori reali, ma riaprono il problema di quale ruolo abbia il modello decisionale: se debba solo descrivere lo stato di fatto, o anche prescrivere un approccio efficace alla decisione.

⁵Maurice Allais (1911-2010), fisico ed economista, premio Nobel per l'Economia nel 1988.

⁶Daniel Kahneman (1934-), psicologo e premio Nobel per l'Economia nel 2002.

⁷Amos Tversky (1937-1996), psicologo.

10.6.2 Il paradosso di Ellsberg

Questo paradosso, proposto da D. Ellsberg⁸ nel 1961, riguarda invece un aspetto più fondamentale, cioè la definizione di probabilità usata nel modello. Il paradosso ipotizza che un'urna contenga tre biglie colorate, di cui una rossa, mentre le altre due possono essere nere o gialle, ma non si sa esattamente di quale colore siano. In un primo test si può scegliere tra due lotterie, relative all'estrazione di una biglia dall'urna:

1. nella prima lotteria, si vince 1 euro quando si estrae una biglia rossa;
2. nella seconda lotteria, si vince 1 euro quando si estrae una biglia nera.

La probabilità di vittoria per la prima lotteria è $1/3$, mentre quella per la seconda è $\alpha/3$, dove $\alpha \in \{0, 1, 2\}$ è il numero sconosciuto di biglie nere nell'urna. La maggior parte dei decisori sottoposti a questo test preferisce la prima lotteria. Secondo la teoria di Von Neumann e Morgenstern, questo significa che

$$u(\ell_1) > u(\ell_2) \Rightarrow \frac{1}{3}u(1) + \frac{2}{3}u(0) > \frac{\alpha}{3}u(1) + \left(1 - \frac{\alpha}{3}\right)u(0)$$

In un secondo test, gli stessi decisori, posti di fronte alla stessa urna, possono scegliere tra altre due lotterie:

1. nella prima lotteria, si vince 1 euro quando si estrae una biglia rossa o gialla;
2. nella seconda lotteria, si vince 1 euro quando si estrae una biglia nera o gialla.

La probabilità di vittoria è $1 - \alpha/3$ per la prima lotteria e $2/3$ per la seconda. La maggior parte dei decisori sottoposti al test preferisce la seconda lotteria. Secondo la teoria di Von Neumann e Morgenstern, questo significa che

$$u(\ell'_2) > u(\ell'_1) \Rightarrow \frac{2}{3}u(1) + \frac{1}{3}u(0) > \left(1 - \frac{\alpha}{3}\right)u(1) + \frac{\alpha}{3}u(0)$$

Se sommiamo membro a membro le due disegualanze, però, otteniamo che

$$u(1) + u(0) > u(1) + u(0) \Rightarrow 0 > 0$$

che è ovviamente un assurdo.

La differenza sostanziale sta nella costruzione della probabilità. Le lotterie vincenti in entrambi i test sono più semplici. In esse, il valore della probabilità è dato: in qualsiasi estrazione si vince con una biglia su tre, oppure con due biglie su tre. Le lotterie perdenti, invece, sono quelle in cui la probabilità di vincere non è sempre uguale, ma cambia probabilisticamente ad ogni estrazione, pur avendo un valore atteso identico a quello delle lotterie vincenti. In certe estrazioni, in effetti, il decisore può essere certo di perdere o certo di vincere. Questo rende il problema intermedio fra la programmazione in condizioni di rischio e quella in condizioni di ignoranza, e gli esseri umani tendono a preferire il rischio all'ignoranza, ma in questa situazione gli assiomi di Von Neumann e Morgenstern sono troppo stretti per essere verificati.

⁸Daniel Ellsberg (1931-), economista e attivista per la libertà di stampa (dal Vietnam negli anni '70 a Wikileaks oggi).

(Lez10, da arricchire con esercizi svolti sull'utilita' stocastica)

Capitolo 11

Teoria delle decisioni

La teoria delle decisioni non è un argomento distinto da quelli svolti in precedenza, ma un'estensione dei concetti esposti sinora a problemi e modelli più complessi, sia per le decisioni in condizioni di ignoranza sia per quelle in condizioni di rischio. In particolare, tiene conto del fatto che *le decisioni in ambiente incerto possono essere distribuite nel tempo su più fasi*, e che quindi *alcune variabili di decisione possono essere fissate dopo che il valore di alcune variabili esogene si è rivelato*. Questo modifica l'ipotesi fatta sinora che prima si dovessero fissare tutte le variabili di decisione e poi venisse rivelato il valore di tutte le variabili esogene.

11.1 Albero delle decisioni

L'*albero delle decisioni* è un strumento di risoluzione per problemi di tipo finito, equivalente alla matrice di valutazione, ma più flessibile. Esso introduce una struttura gerarchica sulle variabili di decisione e sulle variabili esogene.

L'idea fondamentale è che il processo decisionale si svolge in t_{\max} fasi successive, rappresentate da un indice temporale t che cresce da 0 a $t_{\max} - 1$. In ciascuna fase t :

- prima, il decisore fissa un sottovettore di variabili decisionali $x^{(t)}$;
- poi, il mondo esterno fissa un sottovettore di variabili esogene $\omega^{(t)}$.

Si può paragonare il problema a un gioco, nel quale alternativamente il decisore sceglie una mossa $x^{(t)}$ e la natura risponde con una contromossa $\omega^{(t)}$. Quando il decisore deve scegliere $x^{(t)}$, non conosce solo i dati del problema, ma anche le proprie mosse e quelle della natura per tutte le fasi precedenti, da 0 a $t - 1$. Quindi, il problema non consiste più semplicemente nel cercare una soluzione, cioè un vettore di numeri reali $x_i^{(t)}$, ma una *strategia*, cioè un vettore di funzioni $x_i^{(t)}(\omega^{(0)}, \dots, \omega^{(t-1)})$, le cui componenti dipendono in generale dal valore delle variabili esogene note al momento della decisione.

L'albero è costituito da livelli in rigoroso ordine cronologico, dal livello 0, che include solo la radice dell'albero, al livello $2t_{\max} + 1$, che include solo foglie. I livelli sono organizzati a coppie:

- i livelli di indice pari ($2t$) corrispondono al primo tempo della fase t , in cui il decisore compie scelte, cioè fissa il valore delle variabili di decisione $x_i^{(t)}$; gli archi uscenti da uno stesso nodo rappresentano tutti i possibili valori di tali variabili di decisione;

- i livelli di indice dispari ($2t+1$) corrispondono al secondo tempo della fase t , in cui parte dello stato di natura si rivela, cioè fissa il valore delle variabili esogene $\omega_j^{(t)}$; gli archi uscenti da uno stesso nodo v rappresentano tutti i possibili valori di tali variabili esogene, cioè eventi casuali mutuamente esclusivi.

L'ordine cronologico dei livelli fa sì che il dispiegarsi delle decisioni e degli eventi esterni si traduca nel percorrere un ramo dell'albero dalla radice ad una delle foglie. Le foglie dell'albero descrivono le situazioni in cui decisioni e stati di natura sono interamente fissati, cioè le configurazioni (x, ω) del sistema. Ai vari nodi e archi dell'albero sono associati i dati del problema:

- ogni foglia (x, ω) è marcata con il relativo impatto $f(x, \omega)$;
- gli archi uscenti dai nodi di livello dispari possono essere marcati con le probabilità $\pi(\omega)$ degli esiti casuali che essi rappresentano (se queste sono note); la somma delle probabilità per gli archi uscenti da ciascun nodo è unitaria.

Nei problemi studiati sinora, il processo decisionale attraversava una sola fase ($t_{\max} = 1$), per cui l'albero corrispondente ha tre livelli: un livello iniziale di decisione, cioè la radice, uno di svelamento degli scenari, cioè i nodi intermedi, e uno finale di valutazione degli impatti, cioè le foglie (vedi Figura 11.1). Per questi problemi, l'albero di decisione equivale perfettamente alla matrice di valutazione, e al vettore delle probabilità, se quest'ultimo è noto.

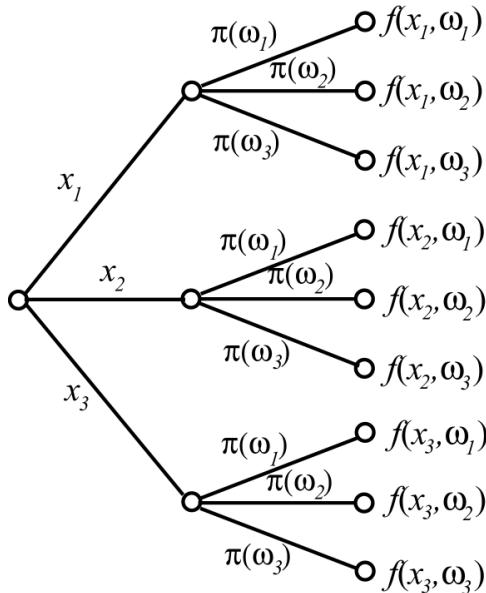


Figura 11.1: Albero di decisione a due livelli per un problema una sola fase di decisione, tre alternative e tre scenari

Per risolvere il problema, si applica il metodo dell'*induzione a ritroso*: si procede cioè a ritroso, dalle foglie alla radice, risalendo l'albero di livello in livello e marcando il nodo padre in base alle marcature dei nodi figli secondo la strategia risolutiva prescelta:

- se il nodo padre corrisponde a una decisione, si marca il padre con la migliore fra le etichette dei nodi figli, a indicare che il decisore sceglie l'alternativa

che produce il risultato migliore; questo corrisponde anche a stabilire il valore delle variabili $x_i^{(t)}$ che descrivono la decisione;

- se il nodo padre corrisponde a un evento esogeno, viene marcato in base al criterio di scelta prescelto per affrontare l'incertezza: caso pessimo, rammarico, equiprobabilità, valore atteso, utilità stocastica, ecc...

Per esempio, se si è deciso di adottare il criterio dell'utilità attesa, la marcatura del nodo padre v deriva da quella dei nodi figli v_ω associati a tutti i possibili esiti dell'insieme Ω_v secondo la relazione:

$$u(v) = \sum_{\omega \in \Omega_v} \pi_\omega u(v_\omega)$$

Adottando altri criteri di scelta, si possono eseguire algoritmi risolutivi diversi sullo stesso albero, e quindi ottenere soluzioni diverse.

Esempio 73 *Un'azienda vuole introdurre sul mercato un nuovo prodotto scegliendo fra tre possibili modelli (A, B, C). Sono stati ipotizzati tre possibili livelli di domanda (Basso, Medio e Alto) e le relative probabilità. I modelli rappresentano le soluzioni $x \in X$ disponibili mentre i livelli di domanda sono i possibili scenari $\omega \in \Omega$. La Tabella 11.1 mostra i profitti $f(x, \omega)$ stimati per ogni configurazione modello/domanda e le probabilità stimate per i possibili scenari. Si vuole massimizzare il profitto totale, applicando i criteri del caso pessimo, di Laplace e del valore atteso.*

$f(x, \omega)$ Modello x	Livello della domanda ω		
	Basso	Medio	Alto
A	200 000	350 000	600 000
B	250 000	350 000	540 000
C	300 000	375 000	490 000
Probabilità π		Basso	Medio
		0.1	0.5
		0.4	

Tabella 11.1: Matrice di valutazione e vettore delle probabilità degli scenari per il lancio di un nuovo prodotto sul mercato

L'albero delle decisioni che rappresenta il problema è riportato in Figura 11.2. Il livello 0 corrisponde alla decisione x e gli archi uscenti dalla radice sono associati ai modelli. Il livello 1 corrisponde allo stato di natura e gli archi uscenti da ogni nodo sono associati ai livelli di domanda, con le relative probabilità $\pi(\omega)$. Si noti che esse si ripetono in ciascuno dei tre sottoalberi associati alle possibili soluzioni. Infatti, le probabilità non dipendono dall'alternativa scelta. Alle foglie sono associati i valori dell'impatto $f(x, \omega)$.

Criterio del caso pessimo *Partendo dalle foglie, marchiamo ciascun nodo padre associato a una soluzione x_i calcolando il valore pessimo $\min_{\omega \in \Omega} f(x_i, \omega)$ fra quelli dei nodi figli, cioè fra gli scenari possibili nel nodo stesso. Questo equivale a calcolare il minimo di ogni riga della matrice di valutazione, come già visto nella Sezione 9.1. Il risultato è riportato sui nodi del livello 1 in Figura 11.3. Al livello 0, invece, poiché esso rappresenta una decisione, si sceglie l'alternativa migliore, che massimizza le marcature dei nodi figli, cioè il modello C. La Figura 11.3 riassume il procedimento.*

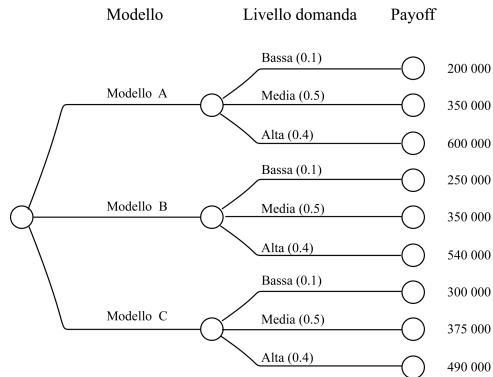


Figura 11.2: Albero delle decisioni per il lancio di un nuovo prodotto sul mercato

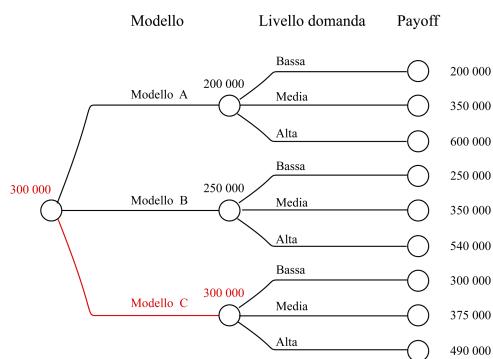


Figura 11.3: Albero delle decisioni per il lancio di un nuovo prodotto sul mercato e risoluzione con il criterio del caso pessimo

Criterio di Laplace Il criterio di Laplace comporta lo stesso procedimento di induzione a ritroso, ma questa volta nei nodi del livello 1 si applica un diverso criterio di marcatura: in ogni nodo si riporta la media aritmetica dei valori dei nodi figli. Al livello 0, invece, poiché esso rappresenta una decisione, si sceglie ancora l'alternativa che massimizza le marcature dei nodi figli, che è ancora il modello C anche se le marcature stesse sono cambiate. La Figura 11.4 riassume il procedimento.

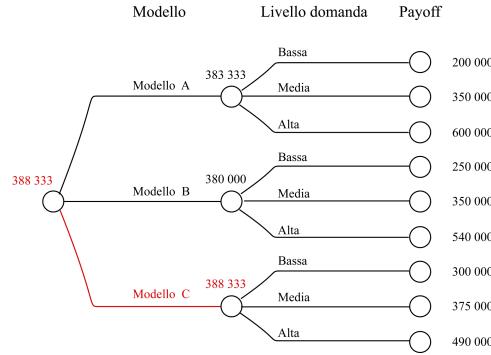


Figura 11.4: Albero delle decisioni per il lancio di un nuovo prodotto sul mercato e risoluzione con il criterio di Laplace

Criterio del valore atteso Partendo dalle foglie, marchiamo ciascun nodo padre associato a una soluzione x_i calcolando il valore atteso $E[f(x_i, \omega)] = \sum_{\omega \in \Omega} \pi_\omega f(x_i, \omega)$. Questo equivale a calcolare il prodotto di ogni riga della matrice di valutazione per il vettore delle probabilità. Questa volta, quindi, si usa l'informazione fornita dal vettore π . Il risultato è riportato sui nodi del livello 1 in Figura 11.5. A questo punto, si può risalire dal livello 1 alla radice, usando la strategia deterministica, dato che la radice corrisponde a una decisione: si sceglie l'alternativa che massimizza le marcature dei nodi figli, cioè il modello A. La marcatura associata viene posta come etichetta del nodo radice ed è la soluzione del problema.

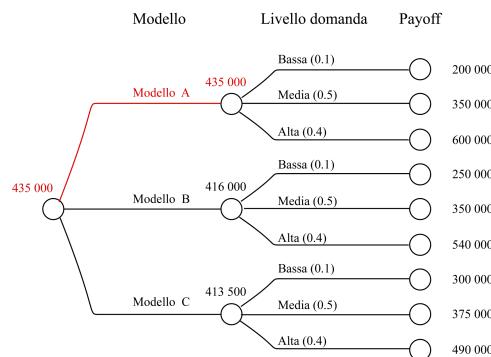


Figura 11.5: Albero delle decisioni per il lancio di un nuovo prodotto sul mercato e risoluzione con il criterio del valore atteso

Albero delle decisioni e matrice di valutazione sono due modi equivalenti di rappresentare i dati di un problema di decisione in ambiente incerto. Il procedimento risolutivo è lo stesso: lavorando sulle righe della matrice o sugli archi dell'albe-

ro, infatti, si eseguono le stesse operazioni. Che vantaggio offre allora l'albero di decisione?

Nel seguito presentiamo tre situazioni nelle quali l'uso dell'albero di decisione si mostra vantaggioso come strumento di modellazione in quanto più flessibile della matrice.

11.2 Scenari condizionati dalla decisione

Quando lo stato di natura è influenzato dalle variabili di decisione, le probabilità degli scenari non costituiscono più un vettore $\pi(\omega)$ di valori assoluti, ma una matrice $\pi(\omega|x)$ di valori condizionati dall'alternativa scelta. Questo modifica solo leggermente il procedimento: nell'albero di decisione ogni nodo ottenuto con una decisione ha archi uscenti associati agli scenari aperti da tale decisione. Bisogna associare a tali archi non le probabilità assolute degli scenari, ma le probabilità condizionate alla decisione presa nel nodo.

Esempio 74 Modifichiamo l'esempio del lancio di un prodotto sul mercato supponendo che i tre scenari di domanda del mercato siano dipendenti dal prodotto lanciato, cioè che, siccome i modelli possibili hanno attrattive diverse tra loro (inconscie, ma stimabili), i tre scenari denominati di bassa, media e alta domanda abbiano probabilità diverse a seconda del modello scelto: un modello più attrattivo aumenta la probabilità di una domanda alta. La Tabella 11.2 riporta le probabilità condizionate $\pi(\omega_j|x_i)$.

x_i	$\pi(\omega_j x_i)$		
	Basso	Medio	Alto
A	0.3	0.5	0.2
B	0.1	0.5	0.4
C	0.1	0.6	0.3

Tabella 11.2: Matrice delle probabilità degli stati di natura condizionate alle alternative per il lancio di un nuovo prodotto sul mercato

La Figura 11.6 riporta l'albero di decisione che corrisponde alla matrice di valutazione originale e alle nuove probabilità condizionate e la risoluzione con il criterio del valore atteso. Si confronti la marcatura degli archi con quella della Figura 11.2: gli archi associati allo stesso scenario in nodi diversi non hanno più probabilità identiche, ma diverse. Questa volta vince il modello B, il quale tende ad avere un'attrattiva superiore per il mercato, e in particolare una probabilità più forte di stimolare una domanda alta.

11.3 Decisioni distribuite in più fasi

Quando le decisioni sono distribuite nel tempo e alcune variabili vengono fissate dopo che lo stato di natura si è parzialmente svelato, l'albero di decisione ha più di due livelli. Questo problema non si riesce a rappresentare in modo elementare su una matrice bidimensionale¹.

Esempio 75 Supponiamo che, una volta lanciato il prodotto si possa decidere se far partire una campagna pubblicitaria di sostegno alle vendite oppure no. In que-

¹Vedremo nella Sezione che è possibile, ma al costo di notevoli complicazioni.

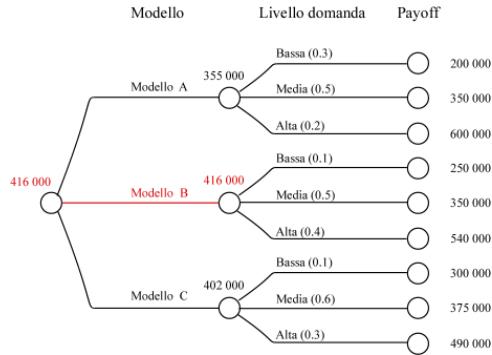


Figura 11.6: Albero delle decisioni per il lancio di un nuovo prodotto sul mercato nel caso di probabilità condizionate

sto caso, le decisioni da prendere sono due: quale prodotto lanciare e se avviare la campagna. La seconda decisione va presa dopo il lancio del prodotto, quando si è misurato se la domanda è bassa, media o alta. A sua volta, la campagna potrebbe avere un esito incerto, e quindi bisognerebbe introdurre un'ulteriore variabile esogena, e un ulteriore livello nell'albero delle decisioni. Per limitare le dimensioni dell'esempio, e per mostrare come si modellano i fenomeni deterministici in questo contesto, supponiamo che l'effetto della campagna sia certo e che sia quello riportato nella Tabella 11.3. I valori di profitto riportato tengono conto da un lato dell'aumento di vendite provocato dalla campagna, dall'altro del costo della campagna stessa, per cui il risultato è talvolta migliore e talaltra peggiore che nella Tabella 11.1.

Modello x	Livello della domanda ω		
	Basso	Medio	Alto
A	220 000	340 000	560 000
B	300 000	380 000	530 000
C	360 000	415 000	480 000

Tabella 11.3: Matrice di valutazione per il lancio di un nuovo prodotto sul mercato dopo il lancio di una campagna pubblicitaria (da confrontare con la Tabella 11.1)

L'ipotesi che la campagna abbia un effetto noto a priori riduce il livello 3 dell'albero di decisione (scenari conseguenti al lancio della campagna) a un insieme di archi singoli ciascuno uscente da un nodo del livello 2 e con probabilità associata unitaria. Se la campagna avesse più esiti potenziali, ciascun nodo emetterebbe più archi, con probabilità di somma unitaria, potenzialmente condizionate sia dalle decisioni sia dagli scenari che si incontrano lungo il cammino dalla radice al nodo padre.

La Figura 11.7 riporta i dati e il procedimento risolutivo per questo problema esteso. Si procede, come sempre, risalendo l'albero dalle foglie alla radice. Il livello 3 è banale e consiste solo nel copiare le marcature delle foglie nei nodi padre. Se alcuni nodi padre avessero più di un arco uscente, si dovrebbe invece scegliere e applicare un criterio per determinarne la marcatura (per esempio, il criterio del valore atteso). La marcatura del livello 2 si ottiene scegliendo il valore massimo fra i figli, dato che si tratta di una decisione (avviare o no la campagna pubblicitaria). La marcatura del livello 1 viene fatta con il criterio del valore atteso, e quella della radice massimizzando. Il risultato, marcato in rosso nella figura, non è una semplice

soluzione, ma una strategia in due fasi, la prima determinata univocamente, la seconda dipendente dallo scenario parziale deciso dalla natura:

1. lanciare il modello C
2. misurare la domanda:
 - se è bassa, avviare la campagna pubblicitaria;
 - se è media, avviare la campagna pubblicitaria;
 - se è alta, non avviare la campagna pubblicitaria.

La strategia ottima prevede non solo la scelta del modello, ma anche l'indicazione di come reagire alla domanda di mercato, una volta che essa si rivela: in caso di domanda alta, non si esegue la campagna, perché il suo costo non è compensato dall'aumento di vendite atteso. Il guadagno atteso dall'applicazione di questa strategia è 439 500 euro.

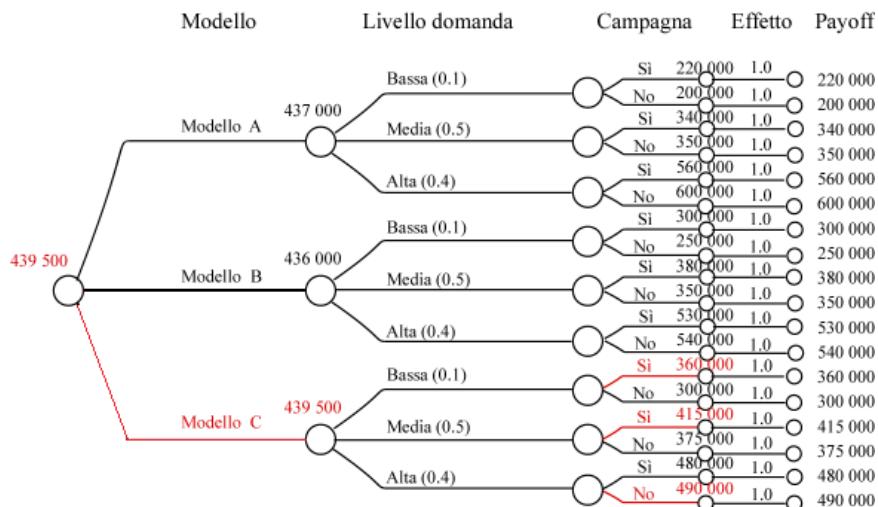


Figura 11.7: Albero delle decisioni per il lancio di un nuovo prodotto sul mercato nel caso di decisione in due fasi

11.4 Esperimenti casuali

Talvolta la stima delle probabilità dei diversi scenari può essere raffinata attraverso l'esecuzione di un “esperimento”. Un esempio tipico riguarda i fenomeni meteorologici. Possiamo stimare la probabilità di pioggia in un determinato periodo sulla base di dati storici. Tuttavia, se consultiamo uno strumento come un barometro, termometro o igrometro (o anche più strumenti insieme), otteniamo delle misure che a loro volta sono correlate in qualche modo con la probabilità di pioggia. Il valore di tali misure è l'esito di un esperimento casuale, e in genere non consente di conoscere con precisione lo stato di natura, cioè di sapere con certezza se pioverà o no. Tuttavia, conoscendo la precisione dello strumento, cioè la correlazione fra gli esiti delle sue misure e i possibili scenari, si può derivare una stima migliore delle probabilità degli scenari stessi.

L'albero di decisione consente di rappresentare tutto questo con un livello aggiuntivo di scenari incerti, questa volta a monte della decisione fondamentale. Siccome bisogna anche decidere se eseguire l'esperimento, occorre aggiungere un altro

livello, relativo a tale decisione, a monte del livello associato agli esiti dell'esperimento. Bisogna infatti tener conto che qualsiasi esperimento in genere ha un costo, e che quindi non è ovvio che convenga eseguirlo: dipende dal vantaggio che se ne ricava in termini di qualità della soluzione scelta. La Figura 11.8 dà una rappresentazione compatta dell'albero di decisione, che illustra solo la struttura a livelli dell'albero stesso, collassando tutti i nodi di un livello in un nodo solo. L'albero ha la seguente struttura:

- il livello 0 si riferisce alla decisione se eseguire o no l'esperimento casuale prima di affrontare il problema dato: la variabile decisionale $x' \in X'$ può assumere solo due valori;
- il livello 1 si riferisce all'esperimento casuale: le variabili esogene $\omega' \in \Omega'$ indicano l'esito dell'esperimento, detto anche *segnalet*;
- il livello 2 si riferisce alla soluzione $x \in X$ scelta per il problema dato;
- il livello 3 si riferisce agli scenari $\omega \in \Omega$ del problema dato.

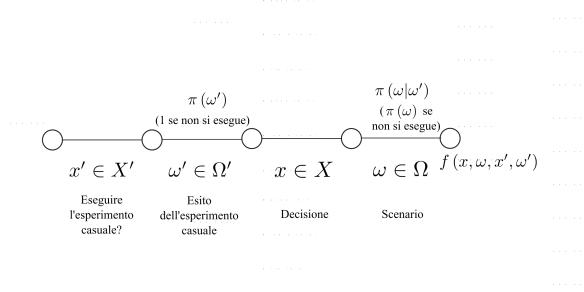


Figura 11.8: Albero delle decisioni in presenza di un esperimento casuale: per semplicità si collassano tutti i nodi dello stesso livello in un nodo solo

Se si vuole applicare il criterio del valore atteso, o quello dell'utilità attesa, bisogna riportare sugli archi relativi agli esiti dell'esperimento, cioè quelli uscenti dai nodi del livello 1, le probabilità $\pi(\omega')$ degli esiti stessi. Sugli archi relativi agli scenari, cioè quelli uscenti dai nodi del livello 3, bisogna invece riportare le probabilità condizionate $\pi(\omega|\omega')$ di tali scenari rispetto agli esiti dell'esperimento. In generale, la probabilità riportata su ogni arco è condizionata da tutto ciò che avviene sugli archi precedenti nel cammino dalla radice. In particolare, le probabilità potrebbero essere anche condizionate dalle decisioni precedenti, come accadeva nella Sezione 11.2, cioè da x' per gli archi uscenti dai nodi del livello 1 e da x' e x per gli archi uscenti dai nodi del livello 3. In questo esempio, però, le decisioni non influiscono sugli scenari. Osserviamo anche che la relazione non è di causa ed effetto, ma di semplice correlazione statistica: le probabilità $\pi(\omega|\omega')$ degli scenari meteorologici sono condizionate dagli esiti della consultazione degli strumenti. Questo non significa che misurare la pressione col barometro modifichi le probabilità di pioggia. Significa che tale misura fornisce informazioni ulteriori rispetto alle frequenze storiche, e tali informazioni alterano la nostra stima della probabilità che piova. Più in generale, l'esperimento casuale non modifica lo stato di natura, ma approfondisce la nostra conoscenza cambiando la stima delle probabilità degli scenari. D'altra parte, l'esperimento non dà garanzie assolute sullo scenario; se lo facesse, ridurrebbe gli scenari a uno solo, cancellando del tutto l'incertezza, riducendo l'intero problema a un modello di Programmazione Matematica.

Il sottoalbero associato alla decisione di non eseguire l'esperimento dovrebbe saltare a pié pari il livello stocastico associato agli esiti dell'esperimento, passando direttamente dal livello 0 (decisione sull'esperimento) al livello 2 (decisione relativa al problema dato). Per conservare la struttura a livelli alternati, si introduce anche in questo sottoalbero un livello 1 fittizio, costituito da un nodo che rappresenta un esperimento degenero, con un solo arco uscente, che rappresenta un esito fittizio di probabilità 1. Le probabilità degli archi sul livello 3 di questo sottoalbero, cioè le probabilità degli stati di natura, non sono condizionate dagli esiti dell'esperimento, che non viene eseguito, ma sono le probabilità stimate a priori in base allo storico.

Infine, le foglie dell'albero sono, come sempre, associate agli impatti di ciascuna configurazione finale del sistema. Le configurazioni non sono le classiche coppie (x, ω) , ma includono anche le variabili di decisione e le variabili esogene relative all'esperimento: sono quindi quaterne (y, ω', x, ω) . Infatti, l'impatto include il costo dell'esperimento nelle configurazioni che richiedono di pagare questo costo. L'esito dell'esperimento, in questo esempio, non ha invece influenza sul costo finale, anche se in generale potrebbe averne.

Definizione 45 *Definiamo valore dell'informazione V la differenza fra l'utilità guadagnata eseguendo l'esperimento e quella guadagnata non eseguendolo.*

Il valore dell'informazione misura il miglioramento consentito dall'esperimento stesso, e quindi il massimo costo che il decisore è disposto a pagare per poterlo eseguire. Un esperimento di costo superiore al valore dell'informazione che rivela non deve essere eseguito. Ovviamente, il tutto si può generalizzare alle situazioni in cui sono disponibili diversi tipi di esperimento: ciascuno di loro ha un proprio sottoalbero e si può calcolare il valore dell'informazione ad esso associata per confronto con il sottoalbero associato al non eseguire alcun esperimento.

11.4.1 Calcolo delle probabilità per l'albero delle decisioni

Per risolvere il problema con il solito metodo dell'induzione a ritroso, occorre associare agli archi dei livelli stocastici le relative probabilità. Ognuna di loro è la probabilità condizionata dalle decisioni e dagli esiti associati a tutti gli archi del cammino che porta dalla radice al nodo padre. Nel caso che stiamo analizzando bisogna riportare sull'albero:

1. al livello 1 le probabilità totali $\pi(\omega')$ degli esiti dell'esperimento;
2. al livello 3 le probabilità $\pi(\omega|\omega')$ degli stati di natura condizionate agli esiti dell'esperimento.

Per fissare le idee, consideriamo l'esempio delle previsioni del tempo, in cui le variabili esogene ω descrivono il tempo futuro e le variabili esogene ω' descrivono la misura di uno strumento. Poiché la misura precede il manifestarsi del tempo, il livello delle variabili ω' precede quello delle variabili ω . L'albero delle decisioni richiede sul livello 1 la probabilità $\pi(\omega')$ di ciascuna misura, e sul livello 3 la probabilità condizionata $\pi(\omega|\omega')$ di ciascun possibile tempo, dato che si sia ottenuta ciascuna possibile misura. Entrambe queste informazioni non sono disponibili. Tuttavia, si possono ricostruire a partire da informazioni disponibili. Ora vediamo quali informazioni sono disponibili e come fare a ricostruire quelle mancanti.

Oltre alle probabilità totali degli scenari ($\pi(\omega)$), si può supporre di avere delle informazioni sull'affidabilità dell'esperimento, cioè sulla probabilità condizionata $\pi(\omega'|\omega)$ che l'esperimento abbia un dato esito ω' in un determinato scenario ω . Queste probabilità derivano dall'esperienza storica. Nel caso delle previsioni del

tempo, le probabilità condizionate indicano con quale frequenza, per esempio, un barometro indica alta pressione dato che la situazione sia destinata a evolvere verso il bel tempo, oppure verso il brutto tempo. Questa è un'informazione storica: non sappiamo se l'alta pressione indicata dal barometro garantirà veramente bel tempo, ma sappiamo che in passato lo ha fatto con una certa frequenza. Il punto fondamentale è che sull'albero occorre riportare $\pi(\omega')$ e $\pi(\omega|\omega')$, mentre le informazioni che si possiedono sono $\pi(\omega)$ e $\pi(\omega'|\omega)$. Per ottenere le une dalle altre, è sufficiente applicare il *teorema di Bayes*.

Teorema 23 Teorema di Bayes: *Data una famiglia di eventi A_i incompatibili a due a due e un evento B :*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} = \frac{P(B \cap A_i)}{P(B)}$$

Nel nostro caso, gli eventi A_i sono gli stati di natura $\omega \in \Omega$, mentre l'evento B è ciascuno degli esiti ω' dell'esperimento casuale (si riapplica il teorema su ogni esito). Quindi

$$\pi(\omega|\omega') = \frac{\pi(\omega'|\omega)\pi(\omega)}{\sum_{\omega \in \Omega} \pi(\omega'|\omega)\pi(\omega)}$$

Infine, le probabilità totali degli esiti dell'esperimento si ottengono a partire da quelle condizionate $\pi(\omega'|\omega)$ e da quelle degli scenari $\pi(\omega)$, sommando i loro prodotti:

$$\pi(\omega') = \sum_{\omega \in \Omega} \pi(\omega', \omega) = \sum_{\omega \in \Omega} \pi(\omega'|\omega)\pi(\omega)$$

Esempio 76 Un turista in vacanza viene informato che nei giorni piovosi la pioggia inizia a mezza mattina e prosegue per l'intero giorno. Gli si raccomanda di guardare il barometro prima di uscire per decidere come vestirsi. Il barometro può indicare bel tempo, variabile e brutto tempo e il tempo può essere bello e asciutto o brutto e piovoso. Inoltre, il turista può scegliere tre tipi di abbigliamento: leggero, leggero ma con l'ombrellino o abbigliamento da pioggia con ombrello, impermeabile, berretto e stivali.

Riassumendo, vi sono due possibili stati di natura ($|\Omega| = 2$):

- $\omega_1 = \text{bel tempo}$
- $\omega_2 = \text{pioggia}$

Le alternative disponibili sono tre ($|X| = 3$):

- $x_1 = \text{abbigliamento leggero}$
- $x_2 = \text{abbigliamento leggero con ombrello}$
- $x_3 = \text{abbigliamento da pioggia}$

L'osservazione del barometro è un esperimento con tre possibili risultati ($|\Omega'| = 3$):

- $\omega'_1 = \text{bel tempo}$
- $\omega'_2 = \text{variabile}$
- $\omega'_3 = \text{pioggia}$

$f(x, \omega)$	ω_1	ω_2
x_1	0	5
x_2	1	3
x_3	3	2

Tabella 11.4: Funzione di costo per il problema meteorologico

ω	ω_1	ω_2
$\pi(\omega)$	0.40	0.60

Tabella 11.5: Probabilità associata agli stati di natura per il problema meteorologico

La Tabella 11.4 riporta i valori del “costo” $f(x, \omega)$ per ogni combinazione di decisione (vestiario) e scenario (tempo). Il problema è quindi di minimizzazione.

Lo storico fornisce la probabilità a priori $\pi(\omega)$ di ogni stato di natura (Tabella 11.5) e l'affidabilità del barometro, cioè la probabilità condizionata $\pi(\omega'|\omega)$ di avere una certa indicazione dal barometro per un certo stato di natura (Tabella 11.6).

$\pi(\omega' \omega)$	ω_1	ω_2
ω'_1	0.60	0.20
ω'_2	0.25	0.30
ω'_3	0.15	0.50

Tabella 11.6: Probabilità dei risultati dell'esperimento condizionate dallo stato di natura per il problema meteorologico

Un barometro “ideale” avrebbe probabilità condizionata pari a 1 per (ω_1, ω'_1) e di (ω_2, ω'_3) e 0 altrove, ma il barometro è ben lungi dall'essere ideale. Tuttavia, l'uso del barometro può migliorare il costo atteso della decisione.

La Figura 11.8 riporta lo schema dell'albero delle decisioni per questo problema:

1. gli archi uscenti dalla radice (primo livello) rappresentano la scelta $x' \in \{0, 1\}$ del decisore se eseguire o no l'esperimento;
2. gli archi del secondo livello rappresentano i risultati ω' dell'esperimento e sono marcati con le loro probabilità totali;
3. gli archi del terzo livello rappresentano la scelta $x(\omega')$ condotta dal decisore in base al risultato dell'esperimento;
4. gli archi del quarto livello rappresentano gli stati di natura ω , sono marcati con le loro probabilità condizionate ai risultati dell'esperimento, e quindi determinano il costo per il decisore, riportato sulle foglie.

Applicando il teorema di Bayes, si procede come segue:

1. si calcolano le probabilità congiunte $p(\omega, \omega')$ (Tabella 11.7), moltiplicando le probabilità condizionate dei risultati per quelle a priori degli stati di natura;
2. se ne ricavano le probabilità a priori dei risultati dell'esperimento $\pi(\omega')$ (Tabella 11.8), sommando sugli stati di natura, cioè riga per riga;
3. da queste si ricavano le probabilità condizionate degli stati di natura rispetto ai risultati $\pi(\omega|\omega')$ (Tabella 11.9), dividendo ogni probabilità congiunta per la probabilità a priori del risultato.

$\pi(\omega' \omega)$	ω_1	ω_2
ω'_1	0.24	0.12
ω'_2	0.10	0.18
ω'_3	0.06	0.30

Tabella 11.7: Probabilità congiunta fra stati di natura e risultati dell'esperimento per il problema meteorologico

ω'	$\pi(\omega')$
ω'_1	0.36
ω'_2	0.28
ω'_3	0.36

Tabella 11.8: Probabilità dei risultati dell'esperimento per il problema meteorologico

$\pi(\omega' \omega)$	ω_1	ω_2
ω'_1	0.66	0.33
ω'_2	0.36	0.64
ω'_3	0.17	0.83

Tabella 11.9: Probabilità degli stati di natura condizionate dai risultati dell'esperimento per il problema meteorologico

Le ultime due famiglie di valori vengono usate per etichettare gli archi dell'albero di decisione.

A questo punto, possiamo risalire l'albero risolvendo il problema con il criterio del valore atteso: nei livelli stocastici, si marca il nodo padre con il valore atteso dei nodi figli; nei livelli deterministici, si marca il nodo padre col valore ottimo (cioè minimo) dei nodi figli. La Figura 11.9 riporta l'albero delle decisioni, con le etichette ottenute a partire dalle foglie risalendo fino alla radice, nell'ipotesi che l'esperimento abbia costo nullo.

I rami in grassetto indicano le scelte del decisore, le quali si configurano come una strategia. Vi sono due strategie ottime, equivalenti fra loro, e consistono nel porre:

$$x' = 1 \text{ (effettuare l'esperimento)} \quad x = \begin{cases} x_1 & \text{se } \omega' = \omega'_1 \\ x_2 & \text{se } \omega' = \omega'_2 \\ x_3 & \text{se } \omega' = \omega'_3 \end{cases}$$

oppure

$$x' = 1 \text{ (effettuare l'esperimento)} \quad x = \begin{cases} x_2 & \text{se } \omega' \in \{\omega'_1, \omega'_2\} \\ x_3 & \text{se } \omega' = \omega'_3 \end{cases}$$

Se invece non si effettuasse l'esperimento, la soluzione migliore sarebbe $x = x_2$. Questo implica che il valore dell'informazione è

$$V = 2.2 - 2.02 = 0.18$$

Se l'esperimento comporta un costo $C > 0$, si deve tenerne conto nel procedimento: le etichette delle foglie del sottoalbero che deriva dalla scelta di fare l'esperimento vanno incrementate di C , per cui gli ultimi tre nodi del terzo livello hanno etichetta incrementata di C , e lo stesso avviene per il secondo nodo del secondo livello. I due nodi al livello 1 hanno etichette 2.2 (senza esperimento) e $2.02 + C$ (con esperimento). L'etichetta della radice sarebbe il minimo dei due valori. Quindi, se

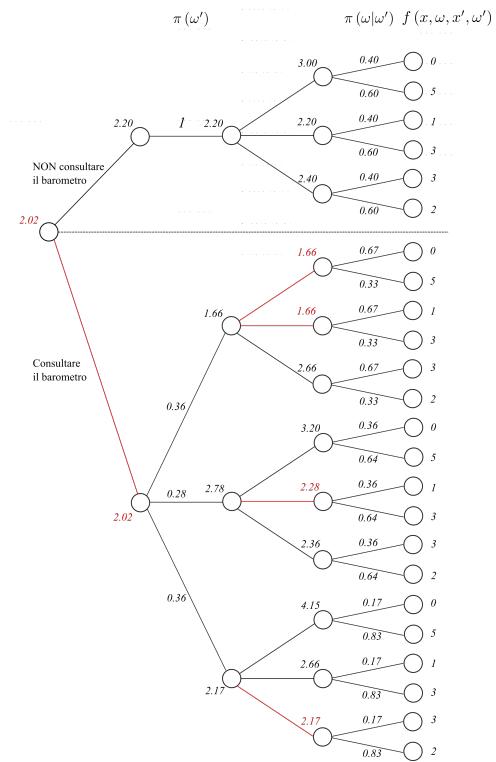


Figura 11.9: Albero delle decisioni per il problema meteorologico

l'esperimento ha un costo $C > 0.18$, non conviene eseguirlo: in tal caso, la strategia ottima è non effettuare l'esperimento e scegliere direttamente x_2 .

11.5 Exercises* † ‡

Esercizio 1

Un problema decisionale ammette tre alternative e due possibili stati di natura. La tabella che segue riporta i benefici associati alle varie configurazioni.

u	ω_1	ω_2
x_1	60	20
x_2	160	-70
x_3	36	46

Si indichi l'alternativa scelta e il valore ad essa associato utilizzando le logiche di scelta di Wald (pessimismo), Hurwicz (con coefficiente di pessimismo $\alpha = 0.4$), Savage (rammarico) e Laplace (ragioni insufficienti).

Ora si supponga che le probabilità degli stati di natura siano $\pi = [0.4 \ 0.6]$ e si indichi l'alternativa scelta in base al criterio del valore atteso.

Infine, si supponga di poter eseguire un esperimento casuale, che ha due esiti possibili (y_1 e y_2) e fornisce informazioni sullo stato di natura in base alla seguente tabella di probabilità condizionate.

$p(y \omega)$	ω_1	ω_2
y_1	0.90	0.20
y_2	0.10	0.80

Si indichi la strategia migliore con il criterio del valore atteso e il valore dell'informazione.

Soluzione

La tabella seguente riporta i valori delle diverse logiche di scelta per le tre alternative.

X	$u_{\text{Wald}}(x)$	$u_{\text{Hurwicz}}(x)$	$u_{\text{Savage}}(x)$	$u_{\text{Laplace}}(x)$
x_1	20	44	100	40
x_2	-70	68	116	45
x_3	46	42	124	41

da cui risulta che l'alternativa migliore per il criterio di Wald è x_3 , per il criterio di Hurwicz è x_2 , per il criterio di Savage è x_1 e per il criterio di Laplace è x_2 .

Introducendo le probabilità, il criterio del valore atteso fornisce: $E[f(x)] = [36 \ 22 \ 42]'$, per cui l'alternativa migliore è x_3 .

Nel caso dell'esperimento, ci sono $3^2 = 9$ strategie possibili, con le seguenti prestazioni.

*The solutions of these exercises have not yet been revised: error reports are welcome.

†I owe several of these exercises to exam texts of professor Alberto Colomni.

‡Tutti gli esercizi di teoria delle decisioni con esperimenti casuali si possono risolvere anche come problemi di programmazione in condizioni di rischio (basta ignorare la possibilità dell'esperimento) o come problemi di programmazione in condizioni di ignoranza (basta ignorare le probabilità degli scenari e applicare le diverse logiche di scelta).

$f(x(y), \omega)$	ω_1	ω_2	u_{media}
(x_1, x_1)	60	20	36
(x_1, x_2)	70	-52	-3.2
(x_1, x_3)	57.6	40.8	47.5
(x_2, x_1)	150	2	61.2
(x_2, x_2)	160	-70	22
(x_2, x_3)	147.6	22.8	72.7
(x_3, x_1)	38.4	25.2	30.5
(x_3, x_2)	48.4	-46.8	-8.7
(x_3, x_3)	36	46	42

Quindi la strategia migliore è la sesta ((x_2, x_3)), con un valore atteso di 72.7. Il valore dell'informazione fornita dall'esperimento casuale è $V = 72.7 - 42 = 30.7$.

Esercizio 2

Si consideri il seguente problema di teoria delle decisioni. Ci sono 10 urne, 8 delle quali di un tipo e 2 di un altro tipo: le prime contengono 4 gettoni rossi e 6 neri, le seconde 9 gettoni rossi e 1 nero. Una di queste viene scelta a caso e posta davanti al decisore, dando luogo a due scenari ω_1 e ω_2 . Il decisore deve cercare di stabilire se l'urna è del primo o del secondo tipo: può scegliere uno dei due tipi (alternativa x_1 o x_2) o astenersi (alternativa x_3). La tabella dei risultati $u(x, \omega)$ è la seguente.

u	ω_1	ω_2
x_1	40	-20
x_2	-5	100
x_3	0	0

Il decisore può, prima di decidere, compiere un esperimento a scelta fra:

1. esperimento e_1 : estrarre un gettone dall'urna, pagando 8;
2. esperimento e_2 : estrarre due gettoni dall'urna, pagando 12.

Ovviamente può anche non fare esperimenti (e_0).

Si formuli e si risolva mediante l'albero delle decisioni il problema di massimizzare il valore atteso della vincita, date le seguenti tabelle di probabilità congiunte fra risultati dell'esperimento e stati di natura per i due esperimenti.

e_1	ω_1	ω_2	e_2	ω_1	ω_2
R	0.32	0.18	RR	0.128	0.162
N	0.48	0.02	RN	0.384	0.036
			NN	0.288	0.002

Soluzione

L'albero delle decisioni ha la seguente struttura:

	e	y	x	ω_1	u
- 29.5	e_0 28	-	x_1	ω_1	40
			28	ω_2	-20
			-	x_2	ω_1 5
			-	16	ω_2 100
			-	x_3	ω_1 0
			-	0	ω_2 0
	e_1 $35.2-8=27.2$	R 32.8	x_1	ω_1	40
			18.4	ω_2	-20
			-	x_2	ω_1 5
			32.8	ω_2	100
			-	x_3	ω_1 0
	e_2 $41.5-12=29.5$	N 37.6	-	0	ω_2 0
			x_1	ω_1	40
			37.6	ω_2	-20
			-	x_2	ω_1 5
			-0.8	ω_2	100
	e_2 $41.5-12=29.5$	RR 53.8	-	x_3	ω_1 0
			-	0	ω_2 0
			x_1	ω_1	40
			6.4	ω_2	-20
			-	x_2	ω_1 5
	e_2 $41.5-12=29.5$	RN 34.0	-	53.8	ω_2 100
			-	x_3	ω_1 0
			-	0	ω_2 0
			x_1	ω_1	40
			34.0	ω_2	-20
	e_2 $41.5-12=29.5$	NR 40.0	-	x_2	ω_1 5
			-	5.5	ω_2 100
			-	x_3	ω_1 0
			-	0	ω_2 0
			x_1	ω_1	40
			40	ω_2	-20
			-	x_2	ω_1 5
			-	-5	ω_2 100
			-	x_3	ω_1 0
			-	0	ω_2 0

da cui si deduce che conviene fare l'esperimento e_2 :

- se si estrae RR, conviene x_2 ;
- negli altri casi, conviene x_1 .

Il valore dell'informazione per e_1 è $V_1 = 7.2$, ma il suo costo è 8. Invece il valore per e_2 è $V_2 = 13.5$, mentre il costo è 12.

Esercizio 3

Si consideri un problema di teoria delle decisioni con due possibili alternative (x_1 e x_2) e due scenari (ω_1 e ω_2). È obbligatorio eseguire un esperimento casuale, che ha tre possibili esiti (y_1 , y_2 e y_3). Gli impatti $f(x, \omega)$ e le probabilità congiunte di esiti dell'esperimento e scenari $\pi(y, \omega)$ sono riportate nelle tabelle seguenti.

$f(x, \omega)$	ω_1	ω_2	$\pi(y, \omega)$	ω_1	ω_2
x_1	0	16	y_1	0.05	0.25
x_2	10	2	y_2	0.10	0.10
			y_3	0.35	0.15

dove gli impatti f sono benefici.

Si risolva il problema con l'albero delle decisioni, applicando prima il criterio del caso pessimo, poi quello del valore atteso.

Si descriva poi nuovamente il problema enumerando in una matrice di valutazione le possibili strategie $x(y)$ sulle righe e gli scenari ω sulle colonne, determinando se vi sono strategie dominate.

Soluzione

Secondo il criterio del caso pessimo, la strategia migliore è:

- $x(y_1) = x_1$,
- $x(y_2) = x(y_3) = x_2$,

che implica $f(x(y(\omega)), \omega) = 9$ per ogni scenario $\omega \in \Omega$.

Secondo il criterio del valore atteso, la strategia migliore è:

- $x(y_1) = x(y_2) = x_1$,
- $x(y_3) = x_2$,

che implica $f(x(y(\omega_1)), \omega_1) = 7$ e $f(x(y(\omega_2)), \omega_2) = 11.8$. Siccome le probabilità dei due scenari sono $\pi(\omega_1) = \pi(\omega_2) = 0.5$, il valore atteso dell'impatto è $E[f(x(y))] = 9.4$.

Le strategie sono banalmente $2^3 = 8$, dato che si tratta di scegliere un'alternativa x fra le due disponibili per ciascuno dei tre esiti y dell'esperimento. La tabella seguente indica ogni strategia con la terna di alternative scelte in corrispondenza a y_1 , y_2 e y_3 nell'ordine.

$f(x(y), \omega)$	ω_1	ω_2
(x_1, x_1, x_1)	0	16
(x_1, x_1, x_2)	7	11.8
(x_1, x_2, x_1)	2	13.2
(x_1, x_2, x_2)	9	9
(x_2, x_1, x_1)	1	9
(x_2, x_1, x_2)	8	4.8
(x_2, x_2, x_1)	3	6.2
(x_2, x_2, x_2)	10	2

La quinta, sesta e settima strategia $((x_2, x_1, x_1), (x_2, x_1, x_2)$ e (x_2, x_2, x_1)) sono dominate dalla quarta $((x_1, x_2, x_2))$.

Esercizio 4

Si consideri il seguente problema con due possibili soluzioni (x_1 e x_2), tre possibili stati di natura (ω_1 , ω_2 e ω_3), un esperimento casuale con quattro possibili risultati (y_1, y_2, y_3 e y_4) e le seguenti tabelle di utilità $u(x, \omega)$ e probabilità congiunte $\pi(y, \omega)$.

$f(x, \omega)$	ω_1	ω_2	ω_3	$\pi(y, \omega)$	ω_1	ω_2	ω_3
x_1	10	20	4	y_1	0.1	0.0	0.3
x_2	12	3	16	y_2	0.1	0.1	0.0
				y_3	0.1	0.0	0.1
				y_4	0.1	0.1	0.0

Quante sono le possibili strategie?

Si risolva il problema usando l'albero delle decisioni.

Qual è il valore dell'informazione fornita dall'esperimento?

Soluzione

Le strategie possibili sono $2^4 = 16$.

Per costruire l'albero delle decisioni, occorre calcolare le probabilità condizionate $\pi(\omega|y)$, riportate nella seguente tabella.

$\pi(\omega y)$	ω_1	ω_2	ω_3
y_1	0.25	0.00	0.75
y_2	0.50	0.50	0.00
y_3	0.50	0.00	0.50
y_4	0.50	0.50	0.00

L'albero delle decisioni ha la seguente struttura:

	e	y	x	ω_1	u
- 14.8	e_0 11.8	-	x_1	ω_1	10
		11.8	9.6	ω_2	20
		-	-	ω_3	4
		-	x_2	ω_1	12
		-	11.8	ω_2	3
	e_1 14.8	-	-	ω_3	16
		y_1	x_1	ω_1	10
		15.0	5.5	ω_2	20
		-	-	ω_3	4
		-	x_2	ω_1	12
-	y_2 15.0	-	15.0	ω_2	3
		-	-	ω_3	16
		-	x_2	ω_1	12
		-	7.5	ω_2	3
		-	-	ω_3	16
	y_3 14.0	x_1	ω_1	10	
		7.0	ω_2	20	
		-	-	ω_3	4
		-	x_2	ω_1	12
		-	14.0	ω_2	3
-	y_4 15.0	-	-	ω_3	16
		x_1	ω_1	10	
		15.0	ω_2	20	
		-	-	ω_3	4
		-	x_2	ω_1	12
		-	7.5	ω_2	3
		-	-	ω_3	16
		-	-	-	-

che suggerisce di eseguire l'esperimento e poi applicare la strategia $s_{11} = (x_2, x_1, x_2, x_1)$, cioè scegliere x_2 se il risultato dell'esperimento è y_1 oppure y_3 e scegliere x_1 in caso contrario. Il valore atteso dell'utilità è 14.8.

Siccome il valore atteso per la decisione senza esperimento è 11.8, il valore dell'esperimento è $V = 14.8 - 11.8 = 3.0$.

Esercizio 5

È dato un problema di decisione in ambiente stocastico, di tipo discreto. Ci sono due possibili azioni e due stati di natura; inoltre sono possibili l'esperimento e_1 (con due risultati) e l'esperimento e_2 (con tre risultati), oltre alla possibilità e_0 di non fare esperimenti. Le tabelle seguenti rappresentano rispettivamente i benefici, le probabilità congiunte per il primo esperimento, quelle per il secondo.

$f(x, \omega)$	ω_1	ω_2	$\pi(y, \omega)$	ω_1	ω_2	$\pi(t, \omega)$	ω_1	ω_2
x_1	500	300	y_1	0.2	0.3	t_1	0.1	0.3
x_2	0	600	y_2	0.1	0.4	t_2	0.1	0.3
						t_3	0.1	0.1

Si risolva il problema con l'albero delle decisioni.

Si dica inoltre quante e quali sono le strategie possibili in base all'esperimento e_1 e in base all'esperimento e_2 .

Soluzione

Le strategie possibili sono $2^2 = 4$ per e_1 e $2^3 = 8$ per e_2 .

Per costruire l'albero delle decisioni, occorre calcolare le probabilità condizionate $\pi(\omega|y)$, riportate nella seguente tabella.

$\pi(\omega y)$	ω_1	ω_2	$\pi(\omega t)$	ω_1	ω_2
y_1	0.4	0.6	t_1	0.25	0.75
y_2	0.2	0.8	t_2	0.25	0.75

			t_3	0.50	0.50
--	--	--	-------	------	------

L'albero delle decisioni ha la seguente struttura²:

	e	y	x	ω_1	u
- 440	e_0 420	-	x_1	ω_1	500
		420	x_1	ω_2	300
		-	x_2	ω_1	0
		-	x_2	ω_2	600
	e_1 430	y_1	x_1	ω_1	500
		380	x_1	ω_2	300
		-	x_2	ω_1	0
		-	x_2	ω_2	600
		y_2	x_1	ω_1	500
		480	x_1	ω_2	300
		-	x_2	ω_1	0
		-	x_2	ω_2	600
	e_2 440	t_1	x_1	ω_1	500
		450	x_1	ω_2	300
		-	x_2	ω_1	0
		-	x_2	ω_2	600
		t_2	x_1	ω_1	500
		450	x_1	ω_2	300
		-	x_2	ω_1	0
		-	x_2	ω_2	600
	t_3 400	x_1	x_1	ω_1	500
		400	x_1	ω_2	300
		-	x_2	ω_1	0
		-	x_2	ω_2	600

Conviene fare l'esperimento e_2 , ammesso che il suo costo sia ≤ 10 e adottare la strategia $s_7 = (x_2, x_2, x_1)$, che opera la scelta x_2 se il risultato dell'esperimento è t_1 o t_2 e la scelta x_1 altrimenti.

Esercizio 6

Si consideri il problema di teoria delle decisioni caratterizzato dalle utilità $f(x, \omega)$, dalle probabilità assolute $\pi(\omega)$ e dalle probabilità condizionate $\pi(y|\omega)$ riportate nelle seguenti tabelle.

$f(x, \omega)$	ω_1	ω_2	$\pi(\omega)$	ω_1	ω_2	$\pi(y \omega)$	ω_1	ω_2
x_1	100	10				y_1	0.8	0.4
x_2	50	30				y_2	0.2	0.6
x_3	25	90						

²Qualche risultato intermedio manca.

Si risolva il problema utilizzando il criterio della massimazione del valore atteso nell'ipotesi di eseguire l'esperimento casuale con esiti y_1 e y_2 .

Si risolva il problema ignorando l'esperimento casuale, sempre con il criterio del valore atteso.

Si ricavi il valore dell'informazione V associato all'esperimento, motivando la risposta.

Soluzione

Vi sono $3^2 = 9$ strategie possibili. La migliore, secondo il criterio del valore atteso, è $s_3 = (x_1, x_3)$, con valore atteso dell'utilità pari a 71.5.

Ignorando l'esperimento, vi sono solo tre possibili soluzioni, rispettivamente con valore atteso dell'utilità pari a $E[f(x_1)] = 55$, $E[f(x_2)] = 40$ e $E[f(x_3)] = 57.5$, per cui la migliore è x_3 .

Il valore dell'informazione fornita dall'esperimento è $V = 71.5 - 57.5 = 14$, dato che eseguire l'esperimento consente di aumentare il valore atteso dell'utilità proprio di questa differenza.

Esercizio 7

Un problema di ottimizzazione in condizioni di incertezza presenta la seguente funzione di beneficio.

$f(x, \omega)$	ω_1	ω_2
x_1	100	0
x_2	300	-200

e le seguenti probabilità assolute per gli scenari: $\pi(\omega) = [0.4 \ 0.6]'$.

Il decisore può fare direttamente la sua scelta, oppure può effettuare degli esperimenti: la relazione tra i risultati del primo esperimento e_1 e gli stati di natura è data dalla tabella delle probabilità congiunte $\pi(y, \omega)$, quella tra risultati del secondo esperimento e_2 e stati di natura dalla tabella $\pi(t, \omega)$.

$\pi(y, \omega)$	ω_1	ω_2	$\pi(t, \omega)$	ω_1	ω_2
y_1	0.30	0.20	t_1	0.25	0.25
y_2	0.10	0.40	t_2	0.15	0.35

Si risolva con l'albero delle decisioni, decidendo quale esperimento effettuare e quale strategia applicare.

Si dica qual è il valore dell'informazione nei due casi.

Soluzione

L'albero delle decisioni ha la seguente struttura:

	e	y	x	ω_1	u
60	e_0	-	x_1	ω_1	100
		40	40	ω_2	0
		-	x_2	ω_1	300
		-	0	ω_2	-200
	e_1	y_1	x_1	ω_1	100
		100	60	ω_2	0
		-	x_2	ω_1	300
		-	100	ω_2	-200
	e_2	y_2	x_1	ω_1	100
		20	20	ω_2	0
		-	x_2	ω_1	300
		-	-100	ω_2	-200
	t_1	t_1	x_1	ω_1	100
		40	50	ω_2	0
		-	x_2	ω_1	300
		-	50	ω_2	-200
	t_2	t_2	x_1	ω_1	100
		30	30	ω_2	0
		-	x_2	ω_1	300
		-	-50	ω_2	-200

La soluzione ottima consiste nel fare l'esperimento e_1 e adottare la seguente strategia: scegliere x_2 se l'esito dell'esperimento è y_1 , scegliere x_1 se l'esito è x_1 .

I valori dell'informazione nei due casi sono $V_1 = 60 - 40 = 20$ e $V_2 = 40 - 40 = 0$.

Esercizio 8

È dato un problema di ottimizzazione in condizioni di incertezza, i cui benefici sono espressi dalla funzione $f(x, \omega)$ e per il quale è possibile fare un esperimento con probabilità congiunte $\pi(x, \omega)$. Le due funzioni sono descritte dalle tabelle seguenti.

$f(x, \omega)$	ω_1	ω_2	ω_3	$\pi(y, \omega)$	ω_1	ω_2	ω_3
x_1	-5	40	10	y_1	0.30	0.05	0.15
x_2	50	0	-10	y_2	0.06	0.40	0.04

Si risolva il problema mediante un albero delle decisioni, utilizzando il criterio del valore atteso.

Si ricavi il valore dell'informazione ottenuta realizzando l'esperimento, motivando la risposta.

Soluzione

Le probabilità assolute degli scenari sono $\pi(\omega) = [0.36 \ 0.45 \ 0.19]'$. Le probabilità condizionate agli esiti dell'esperimento sono $\pi(\omega|y_1) = [0.60 \ 0.10 \ 0.30]'$ e $\pi(\omega|y_2) = [0.12 \ 0.80 \ 0.08]'$.

L'albero delle decisioni ha la seguente struttura:

	e	y	x	ω_1	u
- 29.6	e_0 18.1	-	x_1 18.1	ω_1	-5
				ω_2	40
				ω_3	10
	e_1 29.6	y_1 27.0	x_2 16.1	ω_1	50
				ω_2	0
				ω_3	-10
	e_1 32.2	y_2 32.2	x_1 4.0	ω_1	-5
				ω_2	40
				ω_3	10
		y_2 5.2	x_2 5.2	ω_1	50
				ω_2	0
				ω_3	-10

La soluzione ottima consiste nel fare l'esperimento e_1 e adottare la seguente strategia: scegliere x_2 se l'esito dell'esperimento è y_1 , scegliere x_1 se l'esito è x_1 .

Il valore dell'informazione è $V = 29.6 - 18.1 = 11.5$.

Parte IV

Modelli con decisori molteplici

Capitolo 12

Teoria dei giochi

I problemi decisionali nei quali vi sono più decisori comprendono due casi estremi principali:

1. la *teoria dei giochi*, che studia le situazioni in cui ciascun decisore ha le proprie specifiche variabili di decisione, che fissa in modo indipendente dagli altri decisori;
2. le *decisioni di gruppo*, che studiano le situazioni in cui i decisori condividono le stesse variabili di decisione e devono coordinarsi per fissarne il valore in modo comune.

Ovviamente, possono esistere casi intermedi, in cui alcune variabili sono oggetto di compromessi fra decisori, mentre altre sono fissate autonomamente, ma i due casi estremi esibiscono gli aspetti caratteristici che vale la pena di approfondire.

Il caso della teoria dei giochi viene modellato definendo:

- un insieme di decisori finito $D = \{1, \dots, |D|\}$;
- una regione ammissibile data dal prodotto cartesiano di regioni ammissibili X_d definite sui decisori:

$$X = X^{(1)} \times \dots \times X^{(|D|)} \Leftrightarrow x = [x^{(1)} \ \dots \ x^{(|D|)}]^T$$

e quindi un vettore x di variabili di decisione composto da $|D|$ sottovettori $x^{(d)}$ associati ai singoli decisori e soggetti a vincoli che coinvolgono solo un sottovettore alla volta;

- un ambiente perfettamente deterministico, cioè un solo scenario:

$$\Omega = \{\omega\} \Leftrightarrow |\Omega| = 1$$

per cui non occorre mai menzionare esplicitamente lo scenario ω nel modello;

- una funzione impatto vettoriale costituita da d funzioni monodimensionali associate ai decisori:

$$f = [f^{(1)} \ \dots \ f^{(|D|)}]^T \text{ con } f^{(d)} : X^{(d)} \rightarrow F^{(d)} \subseteq \mathbb{R} \quad d \in D$$

- una funzione Π che definisce per ogni decisore una relazione di preferenza Π_d , la quale consiste semplicemente nel preferire impatti maggiori:

$$\Pi_d = \left\{ (f^{(d)}, f'^{(d)}) \in F^{(d)} \times F^{(d)} : f^d \geq f'^{(d)} \right\}$$

cioè le componenti $f^{(d)}$ dell'impatto sono benefici da massimizzare (se è più opportuno, possono diventare dei costi da minimizzare).

Un problema di teoria dei giochi si può quindi riassumere nella notazione:

$$\max_{\substack{x^{(d)} \in X^{(d)}}} f^{(d)} = f^{(d)}(x^{(1)}, \dots, x^{(|D|)}) \quad d \in D$$

Il punto fondamentale di questa formalizzazione, che differenzia la teoria dei giochi dal semplice accostamento di $|D|$ problemi di Programmazione Matematica su variabili indipendenti e con funzioni obiettivo diverse, è che l'obiettivo $f^{(d)}$ di ciascun decisore $d \in D$ dipende non solo dalle variabili $x^{(d)}$ controllate dal decisore stesso, ma anche da quelle controllate dagli altri. Questo complica notevolmente la risoluzione del problema, fino al punto di mettere in dubbio, in alcuni casi, che abbia senso parlare di soluzione.

La teoria dei giochi introduce una nomenclatura parzialmente diversa da quella usata finora per i problemi decisionali:

- i decisori si possono anche denominare *giocatori*;
- gli impatti si possono anche denominare *payoff*;
- i sottovettori che descrivono le soluzioni per ciascun decisore si possono anche denominare *strategie pure*: un motivo è che i giochi possono svolgersi in più fasi, e quindi alcune variabili possono dipendere dal valore di altre variabili (dello stesso decisore o di altri) fissate in fasi precedenti di gioco; discuteremo però un altro motivo di questa nomenclatura, che vale anche per giochi a una sola fase;
- l'intero vettore che descrive una soluzione, cioè la combinazione delle strategie giocate da tutti i decisori si può anche denominare *profilo di strategie*;
- quando un gioco si svolge in più fasi, ciascun possibile valore assegnato alle variabili fissate in una fase si definisce *mossa*.

Esistono estensioni della teoria dei giochi ai casi di criteri multipli e di ambiente incerto. Le ignoreremo per semplicità e anche perché spesso consistono essenzialmente nel combinare i concetti introdotti per affrontare separatamente ciascuna delle tre complicazioni dei problemi decisionali complessi: la molteplicità dei criteri, degli scenari e dei decisori.

È possibile classificare i problemi di teoria dei giochi in base a diversi punti di vista. Le relative classificazioni si intersecano, in generale, producendo molti casi e sottocasi. In particolare, è possibile distinguere:

1. *giochi non cooperativi*, nei quali i diversi decisori agiscono in modo del tutto indipendente, solo sulla base dei dati elencati nel modello;
2. *giochi cooperativi*, nei quali sono possibili coalizioni fra decisori, accordi vincolanti, trasferimenti di utilità fra decisori o altre simili complicazioni (queste introducono nuovi elementi nel modello, come variabili di decisione per specificare se proporre o no accordi, se accettarli o no, che ammontare di utilità trasferire fra decisori in ogni situazione, ecc...).

Da un altro punto di vista, è possibile distinguere:

1. *giochi a informazione completa*, nei quali ciascun decisore conosce le soluzioni ammissibili e i *payoff* di tutti i decisori;

2. *giochi a informazione incompleta*, nei quali alcune informazioni sono celate ad alcuni decisori (per esempio, ognuno sa che cosa guadagnerebbe da ogni configurazione, ma non sa che cosa guadagnerebbero gli altri).

Per i giochi nei quali il processo decisionale si svolge in più fasi nel tempo, vale la distinzione fra i *giochi a memoria perfetta*, nei quali ogni decisore conosce tutte le mosse passate, proprie e altrui, e quelli in cui ciò non avviene. Informazione completa e memoria perfetta sono due cose diverse: la prima riguarda i dati, la seconda le mosse dei giocatori. Si noti che, anche quando la memoria è perfetta, ogni decisore conosce solo le mosse passate degli altri decisori, e non quelle che vengono eseguite in simultanea, cioè nella stessa fase di gioco.

Nel seguito considereremo solo giochi non cooperativi, a informazione completa e memoria perfetta (se vi sono più fasi di gioco).

Rappresentazioni dei giochi

Esistono due rappresentazioni principali dei giochi:

1. *forma estesa*, nella quale il gioco è rappresentato con un albero;
2. *forma strategica*, nella quale il gioco è rappresentato con una tabella.

Questa distinzione richiama da vicino le due rappresentazioni dei problemi di decisione in condizioni di incertezza: l'albero delle decisioni e la matrice di valutazione. In effetti, come in quel caso, anche qui la distinzione è semplicemente a livello di rappresentazione: ogni gioco si può rappresentare in entrambi i modi. Tuttavia, molto più dei problemi in condizioni di incertezza, ciascun gioco tende ad avere una rappresentazione naturale, mentre quella alternativa appare forzata e scomoda. Nel seguito discutiamo in maggior dettaglio le due rappresentazioni, mostrando esempi di giochi per le quali una rappresentazione è più naturale dell'altra.

12.1 Giochi in forma estesa

Il gioco viene descritto attraverso un albero, la cui struttura riflette quella del gioco stesso. Ogni livello corrisponde ad una scelta di un giocatore, e gli archi uscenti da un nodo descrivono le sue possibili mosse nella configurazione descritta dal nodo. La sequenza dei livelli corrisponde alla sequenza temporale delle scelte dei giocatori, e quindi dipende dalle regole del gioco. I singoli nodi dell'albero vengono definiti *posizioni di gioco* ed ognuna delle foglie, che descrivono le possibili conclusioni del gioco, è associata al vettore $f = [f^{(1)} \dots f^{(|D|)}]^T$ degli impatti determinati per ogni giocatore dalle scelte $x^{(d)}$ compiute da tutti i giocatori lungo il percorso che collega la radice alla foglia stessa.

I giochi in forma estesa con mosse sequenziali possono essere analizzati con lo stesso approccio degli alberi di decisione. L'unica differenza sta nel fatto che i livelli sono tutti controllati da giocatori (a meno che il gioco non sia stocastico), e quindi non occorre modellare il comportamento della natura con qualche criterio più o meno opinabile (caso pessimo, valore atteso, ecc...). Ad ogni livello, il nodo padre viene deterministicamente marcato con la migliore fra le etichette dei nodi figli, dove "migliore" va riferito alla funzione obiettivo $f^{(d)}$ del giocatore che compie la scelta.

Esempio: il gioco del tris in forma estesa

Un semplice esempio di gioco in forma estesa è il gioco del tris, o filetto, o tic-tac-toe. In una scacchiera 3×3 , due giocatori inseriscono alternativamente una croce o un cerchio. Vince il giocatore che completa una sequenza di tre simboli uguali, in verticale, orizzontale o diagonale. Se la scacchiera è piena e non vi sono sequenze vincenti, la partita è pari. La Figura 12.1 illustra parte dell'albero di gioco del tris.

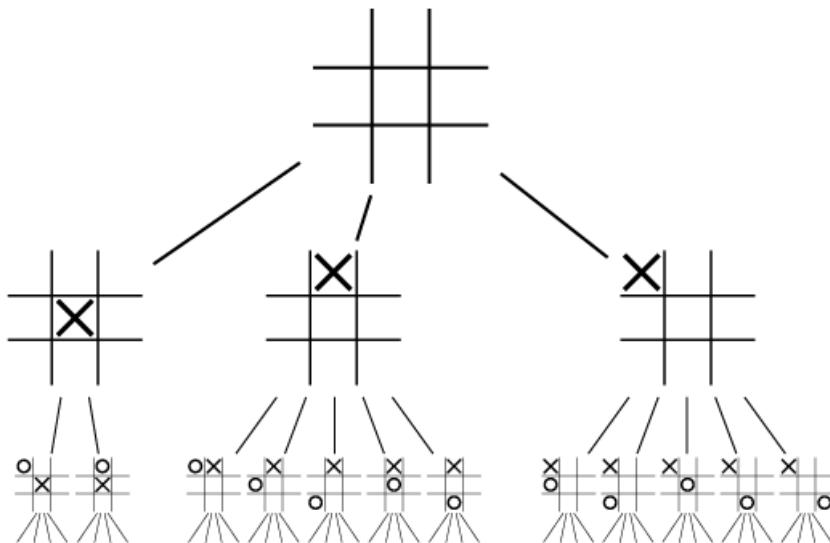


Figura 12.1: Albero del gioco del tris

È possibile analizzare l'albero di gioco con il metodo dell'*induzione a ritroso* già presentato nella Sezione 11.1 e proposto da Von Neumann e Morgenstern. Risalendo dalle posizioni finali di gioco con i relativi impatti (vittoria per la croce, vittoria per il cerchio o patta), si marca ogni nodo con l'etichetta migliore fra quelle dei nodi figli, dove “migliore” si riferisce all’impatto del giocatore associato al livello stesso. Quindi, se in un livello associato al giocatore con il cerchio il nodo padre si trova di fronte anche solo a un nodo figlio marcato con la vittoria del cerchio, sarà marcato anch’esso allo stesso modo; altrimenti, se vi è anche solo un nodo figlio marcato con esito pari, sarà marcato pari; infine, se tutti i nodi figli portano alla vittoria della croce, sarà marcato con la croce. Per i livelli riferiti al giocatore con la croce, valgono le regole complementari. Nel caso del tris, il risultato è che la radice risulta marcata come patta, cioè il tris, giocato razionalmente, conduce sempre alla parità.

Esempio: il gioco di Marienbad in forma estesa

Sono dati diversi mucchi di fiammiferi, ciascuno composto di n_i fiammiferi. Ad ogni turno, un giocatore preleva un numero qualsiasi di fiammiferi da uno solo dei mucchi. Perde il giocatore che preleva l’ultimo fiammifero. La Figura 12.2 illustra l’albero del gioco nel caso molto semplificato in cui ci sono due mucchi e $n_1 = n_2 = 2$. L’albero è stato semplificato considerando che, se due nodi sullo stesso livello corrispondono a situazioni “simmetriche” (cioè in cui la cardinalità dei due sottoinsiemi si scambia), il seguito del gioco, e in particolare il risultato, è necessariamente identico, salvo scambiare le mosse:

- dalla radice, oltre agli archi $(1, 2)$ e $(1, 3)$ uscirebbero altri due archi diretti a nodi “simmetrici” dei primi due, nei quali uno o due fiammiferi sono prelevati dal secondo mucchio;
- dal nodo 5, oltre all’arco $(5, 11)$ uscirebbe un altro arco diretto a un nodo nel quale un fiammifero è prelevato dal secondo mucchio.

Questa semplificazione fa sì che l’albero non sia nemmeno più un albero, ma un grafo aciclico orientato a livelli. D’altra parte, si tratta solo di una rappresentazione compatta di un albero, che riduce il numero dei nodi, e quindi il tempo necessario a studiare e risolvere il gioco. La riduzione è simile a quella ottenuta dai metodi di programmazione dinamica per i problemi di ottimizzazione, anche se la situazione è molto diversa.

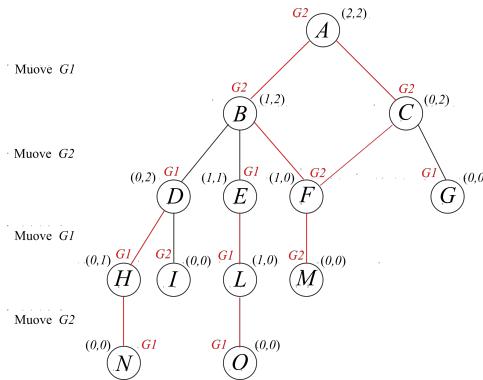


Figura 12.2: Albero del gioco dei fiammiferi (compattato accorpando le posizioni di gioco equivalenti) e soluzione con l’induzione a ritroso: ogni nodo ha un’etichetta alfabetica ed è marcato con la posizione di gioco corrente (n_1, n_2) ; gli archi rappresentano le mosse possibili; quelli marcati in rosso indicano le strategie ottime per i due giocatori (nel nodo radice A il giocatore $G1$ ha due strategie “ottime”). Se entrambi i giocatori giocano al meglio, vince $G2$.

La Figura 12.2 applica il procedimento di risoluzione che risale l’albero, mostrando che, se i giocatori sono razionali, il secondo vince. Il giocatore $G1$ ha due strategie ottime. Entrambe lo portano a perdere, ma sono le migliori possibili (they allow to win if $G2$ makes a mistake). Indicando le singole mosse con il nome degli archi:

Strategia	Posizione			
	A	D	E	F
$x^{(1)}$	(A, B)	(D, H)	(E, L)	(F, M)
$x'^{(1)}$	(A, C)	(D, H)	(E, L)	(F, M)

Il giocatore $G2$ ha una strategia ottima, che lo porta a vincere:

Strategia	Posizione			
	B	C	H	L
$x^{(2)}$	(B, F)	(C, F)	(H, N)	(L, O)

Esempio: il gioco della morra cinese in forma estesa

Un altro esempio di gioco in forma estesa è il gioco della morra cinese, in cui i giocatori *simultaneamente* scelgono un simbolo fra carta, forbice e sasso. La Figura 12.3 illustra l’albero di gioco della morra cinese. La differenza fondamentale

rispetto al gioco precedente è che le mosse sono simultanee, per cui l'ordine fra i due livelli è arbitrario e il secondo giocatore non ha informazione perfetta, cioè non sa in quale ramo dell'albero si trova. Nella figura, questo viene rappresentato dalle linee tratteggiate che collegano i nodi B , C e D . Questa complicazione impedisce al gioco di essere risolto con l'induzione a ritroso, dato che la strategia da essa suggerita fornisce la mossa da compiere in funzione della posizione esatta in cui ci si trova e richiede tre mosse diverse per i tre nodi dell'insieme $\{B, C, D\}$. Per poter risolvere il gioco in forma estesa, occorre rappresentare l'incompletezza dell'informazione. Questo avviene partizionando l'albero di gioco in sottoinsiemi di nodi tutti appartenenti allo stesso livello. Il giocatore associato a quel livello sa che si trova nel sottoinsieme, ma non in quale nodo. Di conseguenza, le sue strategie vanno espresse non come funzioni che indicano una mossa da eseguire in ogni nodo, ma come funzioni che indicano una mossa da eseguire in ogni sottoinsieme.

Definizione 46 Si dice insieme informativo (o information set) ciascun sottoinsieme di nodi in un livello di un albero di gioco fra i quali il giocatore associato al livello non è in grado di discriminare.

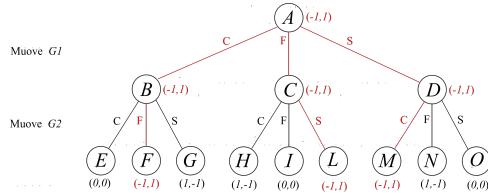


Figura 12.3: Albero del gioco della morra cinese: apparentemente, vince sempre il giocatore G_2 , che però in realtà non è in grado di applicare la strategia vincente, perché non sa se si trova nello stato B , C o D ; per lo stesso motivo, se si scambiano i livelli di gioco, appare ingannevolmente vincitore il giocatore G_1 .

12.2 Giochi in forma strategica

I giochi in forma strategica vengono rappresentati attraverso una matrice $|D|$ -dimensionale, che riporta in ogni casella i *payoff* di ciascun giocatore per il profilo di strategie (cioè la combinazione di strategie giocate dai singoli giocatori) che individua la casella. Se i giocatori sono due, uno si chiama *giocatore di riga* e le sue strategie corrispondono alle righe della matrice, l'altro *giocatore di colonna* e le sue strategie corrispondono alle colonne.

Esempio: il gioco della morra cinese in forma strategica

La Tabella 12.1 riporta il gioco della morra cinese in forma strategica. Vi sono due giocatori, e ognuno ha tre strategie pure a disposizione: carta, forbici e sasso. La carta vince contro il sasso, il sasso contro le forbici e le forbici contro la carta.

Il *payoff* di ciascuno è 1 nel caso di vittoria, 0 in caso di pareggio e -1 in caso di sconfitta. Ad esempio, se il primo giocatore gioca “carta” e il secondo “sasso”, il *payoff* del primo giocatore è pari a 1 e quello del secondo a -1 .

Questa rappresentazione rende evidente il fatto che non vi sono strategie vincenti, dato che nessuna riga o colonna è contenuta solo in impatti vincenti per uno dei due giocatori.

	C	F	S
C	(0, 0)	(-1, 1)	(1, -1)
F	(1, -1)	(0, 0)	(-1, 1)
S	(-1, 1)	(1, -1)	(0, 0)

Tabella 12.1: Tabella dei *payoff* per il gioco della morra cinese**Esempio: il gioco dei fiammiferi in forma strategica**

Per costruire la tabella dei payoff, bisogna aver ben chiaro a che cosa corrispondano le righe e le colonne della matrice: esse rappresentano strategie, cioè non singole scelte, ma funzioni che suggeriscono a un decisore che mossa compiere in qualsiasi situazione in cui deve prendere una decisione, ovvero funzioni per ogni decisore che associano ad ogni nodo legato a quel decisore un arco uscente da tale nodo. Una strategia si può quindi definire in termini di sottoinsiemi di archi dell'albero di gioco.

Definizione 47 *Una strategia per il giocatore $d \in D$ corrisponde a un sottoinsieme di archi dell'albero di gioco che sia*

- costituito da archi associati al giocatore d ,
- coerente, cioè non contenga più archi uscenti dallo stesso nodo,
- massimale, cioè che contenga un arco per ogni nodo dotato di archi uscenti.

Una strategia per un giocatore, infatti, descrive mosse del giocatore stesso, prescrive una sola mossa da eseguire e fornisce indicazioni in ogni situazione possibile. Per esempio, riferendosi all'albero del gioco dei fiammiferi rappresentato nella Figura 12.2, l'arco (2, 6) non può far parte di una strategia per il giocatore A , dato che descrive una mossa del giocatore B . E ancora, gli archi (1, 2) e (1, 3) non sono coerenti fra loro, dato che indicano mosse alternative a partire dalla posizione 1. Infine, il sottoinsieme $\{(1, 2), (5, 11)\}$ non è una strategia completa, dato che non specifica che cosa fare se ci si trova nel nodo 4.

Va osservato che questa definizione di strategia tende ad essere ridondante, dato che i sottoinsiemi contengono anche:

- scelte forzate (per esempio, l'arco (5, 11) è l'unico a uscire dal nodo 5, per cui appare ridondante specificarlo in qualsiasi strategia);
- scelte impossibili (per esempio, la strategia $\{(1, 3), (4, 10), (5, 11), (6, 12), (7, 13)\}$ contiene gli archi (4, 10), (5, 11) e (6, 12), che non sarà mai possibile scegliere una volta eseguita la mossa (1, 3).

Ora elenchiamo le strategie per il gioco dei fiammiferi. Per il giocatore A , vi sono cinque nodi associati a decisioni: $\{1, 4, 5, 6, 7\}$. Il nodo 8, infatti, giace su un livello associato al giocatore A , ma non ha archi uscenti: è una configurazione finale, non associata a decisioni. Le possibili strategie per il giocatore A sono i sottoinsiemi massimali e coerenti di archi associati ad A , cioè:

A1 $\{(1, 2), (4, 9), (5, 11), (6, 12), (7, 13)\}$,

A2 $\{(1, 2), (4, 10), (5, 11), (6, 12), (7, 13)\}$,

A3 $\{(1, 3), (4, 9), (5, 11), (6, 12), (7, 13)\}$,

A4 $\{(1, 3), (4, 10), (5, 11), (6, 12), (7, 13)\}$.

Come si può vedere, tre archi compaiono ovunque, dato che descrivono scelte forzate. Nel seguito, li cancelleremo per semplicità. D'altra parte, le strategie $\{(1, 3), (4, 9)\}$ e $\{(1, 3), (4, 10)\}$ differiscono per un arco che rappresenta una scelta impossibile. Quindi, le due strategie sono in effetti identiche. Le accorperemo in una sola strategia, rappresentata per semplicità con $\{(1, 3)\}$. Il giocatore B prende decisioni nei nodi $\{2, 3, 9, 11\}$, dato che i nodi 10, 12 e 13 rappresentano configurazioni finali. Ne derivano le seguenti strategie:

B1 $\{(2, 4), (3, 7), (9, 14), (11, 15)\}$,

B2 $\{(2, 4), (3, 8), (9, 14), (11, 15)\}$,

B3 $\{(2, 5), (3, 7), (9, 14), (11, 15)\}$,

B4 $\{(2, 5), (3, 8), (9, 14), (11, 15)\}$,

B5 $\{(2, 6), (3, 7), (9, 14), (11, 15)\}$,

B6 $\{(2, 6), (3, 8), (9, 14), (11, 15)\}$.

Per semplicità, nel seguito cancelleremo gli archi $(9, 14)$ e $(11, 15)$, che descrivono scelte forzate e compaiono in tutte le strategie. La matrice dei payoff è descritta nella Tabella 12.2. Come si vede, il giocatore di colonna può vincere sempre con la strategia $\{(2, 6), (3, 7)\}$, qualunque strategia adotti il giocatore di riga. Il giocatore di riga, invece, non ha questa possibilità: tutte le sue strategie possono portare alla vittoria o alla sconfitta secondo le scelte dell'avversario.

	$\{(2, 4), (3, 7)\}$	$\{(2, 4), (3, 8)\}$	$\{(2, 5), (3, 7)\}$	$\{(2, 5), (3, 8)\}$	$\{(2, 6), (3, 7)\}$	$\{(2, 6), (3, 8)\}$
$\{(1, 2), (4, 9)\}$	(1,-1)	(1,-1)	(1,-1)	(1,-1)	(-1,1)	(-1,1)
$\{(1, 2), (4, 10)\}$	(-1,1)	(-1,1)	(1,-1)	(1,-1)	(-1,1)	(-1,1)
$\{(1, 3)\}$	(-1,1)	(1,-1)	(-1,1)	(1,-1)	(-1,1)	(1,-1)

Tabella 12.2: Matrice dei payoff per il gioco dei fiammiferi in forma strategica

12.2.1 Dominanza fra strategie

Definizione 48 Date due strategie $x^{(d)}$ e $x'^{(d)}$ per il giocatore d , si dice che $x^{(d)}$ domina $x'^{(d)}$ quando

$$f^{(d)}(x^{(1)}, \dots, x^{(d)}, \dots, x^{(|D|)}) \geq f^{(d)}(x^{(1)}, \dots, x'^{(d)}, \dots, x^{(|D|)})$$

per ogni $x^{(j)} \in X^{(j)}, j \in D \setminus \{d\}$.

Questo significa che, per qualsiasi comportamento degli altri giocatori $j \neq d$, la strategia $x^{(d)}$ produce per il giocatore d un impatto migliore della strategia $x'^{(d)}$. Poiché si assume che tutti i giocatori siano razionali, cioè abbiano una relazione di preferenza che massimizza il proprio *payoff*, la strategia dominata non verrà mai scelta.

Esempio 77 Consideriamo la seguente matrice di payoff per un gioco a $|D| = 2$ giocatori.

	1	2	3
1	(4,5)	(5,0)	(5,2)
2	(2,6)	(9,1)	(3,2)
3	(3,2)	(2,8)	(6,0)

Rispetto al giocatore di colonna, la terza colonna è evidentemente una strategia dominata dalla prima. Infatti, offre payoff pari a 2, 2 e 0 a seconda della strategia scelta dal primo giocatore, e ciascuno di questi payoff è peggiore di quelli offerti dalla prima colonna (5, 6 e 2). Entrambi i giocatori lo sanno, e quindi possono rimuovere la colonna dominata.

	1	2
1	(4,5)	(5,0)
2	(2,6)	(9,1)
3	(3,2)	(2,8)

Analogamente, la terza riga è dominata dalla prima per il primo giocatore (payoff pari a 3 e 2, contro 4 e 5), per cui si può togliere. Si noti che in partenza tale dominanza non esisteva, perché la terza colonna conteneva un guadagno pari a 6 contro 5. Ma la strategia va eliminata comunque, perché risulta vincente solo nell'ipotesi che l'avversario giochi con una strategia dominata, cosa che possiamo escludere a priori. Si potrebbe parlare di una “dominanza nascosta”.

	1	2
1	(4,5)	(5,0)
2	(2,6)	(9,1)

A questo punto, la seconda colonna è dominata dalla prima (anche questa dominanza era nascosta in precedenza), e quindi si può togliere: il giocatore di colonna non può razionalmente applicare che la prima strategia.

	1
1	(4,5)
2	(2,6)

Ciò significa che il primo giocatore applicherà la prima strategia:

	1
1	(4,5)

L'eliminazione successiva di strategie dominate è un metodo per semplificare il gioco: non sempre porta a una sola strategia per entrambi i giocatori.

Esempio: dominanza in giochi continui

Consideriamo un esempio di gioco con insiemi ammissibili continui: anche se diventa impossibile la rappresentazione matriciale, i concetti di base rimangono validi. Supponiamo che due persone abbiano idee diverse sulla temperatura ideale di una stanza: il giocatore 1 preferirebbe una temperatura di 22 gradi, il giocatore 2 una temperatura di 20 gradi. Supponiamo che il giocatore 1 possa agire sul termostato, impostando un valore compreso fra 16 e 26 gradi, mentre il giocatore 2 può agire su una finestra, modificando il suo grado di apertura in modo continuo fra gli estremi “finestra chiusa” e “finestra spalancata”. Ovviamente, il modello fisico del modo in cui le due variabili di controllo interagiscono per determinare la temperatura

effettiva della stanza è complicato e coinvolge molti altri dati e grandezze, fra cui alcuni valori incogniti. Semplifichiamo il tutto supponendo che

$$T(x_1, x_2) = (x_1 - 5x_2)$$

dove x_1 corrisponde alla temperatura impostata sul termostato dal giocatore 1 e x_2 corrisponde alla frazione di apertura della finestra decisa dal giocatore 2. Inoltre, supponiamo che per entrambi il disagio sia misurato dal quadrato della differenza fra temperatura effettiva e temperatura desiderata. Il problema diventa:

$$\begin{aligned} \min_{x_1} f^{(1)} &= (T(x_1, x_2) - 22)^2 \\ \min_{x_2} f^{(2)} &= (T(x_1, x_2) - 20)^2 \\ x_1 &\in [16; 26] \\ x_2 &\in [0; 1] \end{aligned}$$

Confrontiamo due strategie per il giocatore 1: $x_1 = 22$ e $x_1 = \bar{x}_1 < 22$. In altre parole, ci stiamo chiedendo se abbia senso per il giocatore 1 impostare temperature più basse di 22 gradi. La risposta intuitiva è che, siccome l'avversario può solo abbassare la temperatura, non ha senso per il giocatore 1 impostare valori inferiori al proprio ottimo. Infatti:

$$T(\bar{x}_1, x_2) < T(22, x_2) \leq 22 \Rightarrow f^{(1)}(\bar{x}_1, x_2) > f^{(1)}(22, x_2)$$

per ogni $x_2 \in X^{(2)} = [0; 1]$. Si noti che le componenti dell'impatto sono costi, e quindi la definizione di strategia dominata è esattamente complementare a quella data nella Definizione 48. Dunque, la strategia \bar{x}_1 è dominata per ogni valore $\bar{x}_1 < 22$. Questo significa che il giocatore 1 non imposterà mai il termostato a valori < 22 gradi e che si può ridurre $X^{(1)} = [22; 26]$.

Corrispondentemente, il giocatore 2 può usare questa informazione per escludere strategie dominate. Possiamo confrontare la strategia $x_2 = 0.4$ e la strategia $x_2 = \bar{x}_2 < 0.4$. Siccome il giocatore 1 imposterà sempre il termostato a valori ≥ 22 gradi, il giocatore 2 ha convenienza ad aprire sempre la finestra abbastanza da far calare la temperatura interna di almeno 2 gradi. Infatti:

$$T(x_1, \bar{x}_2) > T(x_1, 0.4) \geq 20 \Rightarrow f^{(2)}(x_1, \bar{x}_2) > f^{(2)}(x_1, 0.4)$$

per ogni $x_1 \in X^{(1)} = [22; 26]$. Dunque, la strategia \bar{x}_2 è dominata per ogni valore $\bar{x}_2 < 0.4$. Questo significa che il giocatore 2 non aprirà mai la finestra meno di 0.4 e che si può ridurre $X^{(2)} = [0.4; 1]$.

A questo punto, si può tornare ad analizzare le strategie per il giocatore 1, dato che ora egli sa che la finestra non sarà mai chiusa. Quindi, gli conviene impostare una temperatura più alta di quella ottimale a finestra chiusa. Con passaggi analoghi a prima, si può dimostrare che non ha senso impostare temperature inferiori a 24 gradi. Rimpallando sul secondo giocatore, ciò comporta che gli conviene aprire la finestra almeno per un fattore 0.8. Di conseguenza, al primo giocatore conviene portare il termostato alla temperatura massima, cioè 26 gradi. A questo punto, l'unica possibilità rimasta al secondo giocatore è di aprire completamente la finestra. Entrambi i giocatori, quindi, hanno una sola strategia non dominata:

$$x_1^* = 26 \quad x_2^* = 1.0$$

Questa coppia di strategie porta a una temperatura pari a $T = 21$ gradi, che corrisponde a un costo pari a $f^{(1)} = f^{(2)} = 1$ per entrambi i giocatori. Una

cosa interessante è che lo stesso risultato si sarebbe potuto ottenere impostando il termostato a 21 gradi e tenendo la finestra chiusa ($x_1^* = 21$ e $x_2^* = 0.0$). D'altra parte, questa soluzione indurrebbe ciascuno dei due giocatori ad agire sulla propria variabile per migliorare il proprio *payoff*, dato che la temperatura è troppo bassa per il primo e troppo alta per il secondo. La soluzione determinata più sopra ha lo stesso difetto, ma non spinge i giocatori a modificare le proprie scelte dato che ognuno dei due ha già portato la propria variabile al valore estremo.

12.2.2 Equilibrio

Definizione 49 Si definisce un profilo di strategie $(x^{*(1)}, \dots, x^{*|D|}) \in X$ come punto di equilibrio, ovvero equilibrio di Nash, quando

$$f^{(d)}(x^{*(1)}, \dots, x^{*(d)}, \dots, x^{*|D|}) \geq f^{(d)}(x^{*(1)}, \dots, x^{(d)}, \dots, x^{*|D|})$$

per ogni $d \in D, x^{(d)} \in X^{(d)}$.

Questo concetto di equilibrio è stato studiato da Nash¹. Il significato intuitivo della definizione è che ogni giocatore che si allontana dal punto di equilibrio finisce col danneggiare sé stesso, nell'ipotesi che tutti gli altri giocatori scelgano di rimanervi.

Si notino le differenze e le somiglianze con la definizione di strategia dominante: qui si confronta una strategia del giocatore d con tutte le altre (anziché con una sola), ma si assume di fissare una sola combinazione di strategie per gli altri giocatori (anziché di considerarle tutte). Da un lato la definizione è più forte, dall'altro più debole. Siccome un equilibrio non può essere dominato, nel (raro) caso in cui la dominanza consente di cancellare tutte le strategie tranne una, questa è necessariamente di equilibrio. In generale, però, ci saranno più strategie residue, e non tutte saranno di equilibrio.

Metodi per determinare gli equilibri

Nel caso dei giochi finiti descritti in forma strategica, un equilibrio è una casella della matrice dei *payoff*. Per trovare gli equilibri si può usare il *metodo della segnatura dei pagamenti*, che consiste nel:

- marcare in ogni colonna il *payoff* massimo per il giocatore di riga;
- marcare in ogni riga il *payoff* massimo per il giocatore di colonna.

Una cella i cui *payoff* sono tutti marcati è un equilibrio di Nash.

Nell'esempio riportato nella Tabella 12.3, l'unico equilibrio di Nash è la casella (3, 1). Si noti che in questo gioco non vi sono strategie dominate. Se ogni giocatore assume che l'altro non cambi la propria strategia, il profilo di strategie (3, 1) si conserva, dato che è il migliore di tutta la riga rispetto ai *payoff* di colonna e di tutta la colonna rispetto ai *payoff* di riga.

Un metodo del tutto equivalente per trovare gli equilibri di Nash è il *metodo del flusso di frecce*, nel quale si tracciano da ogni casella una freccia diretta alla casella di valore massimo in tutta la colonna rispetto alle utilità di riga e una freccia diretta alla casella di valore massimo in tutta la riga rispetto alle utilità di colonna. Gli equilibri di Nash sono le caselle con sole frecce entranti.

¹John Forbes Nash (1928-2015), matematico statunitense, vincitore del Premio Nobel per l'Economia nel 1994.

	1	2	3
1	(4, $\bar{3}$)	(5, 1)	($\bar{6}$, 2)
2	(2, 1)	(8, $\bar{4}$)	(3, 3)
3	($\bar{5}$, 9)	($\bar{9}$, 6)	(2, 8)

Tabella 12.3: Una matrice di *payoff* con equilibrio di Nash nella posizione (3, 1)

È possibile che un gioco non ammetta equilibri di Nash, ed è possibile che ammetta equilibri multipli.

Esempio 78 Il gioco della morra cinese rappresentato nella Tabella 12.4 con le relative segnature non ha equilibri.

	C	F	S
C	(0, 0)	(-1, $\bar{1}$)	($\bar{1}$, -1)
F	($\bar{1}$, -1)	(0, 0)	(-1, $\bar{1}$)
S	(-1, $\bar{1}$)	($\bar{1}$, -1)	(0, 0)

Tabella 12.4: Tabella dei payoff per il gioco della morra cinese con le segnature: non esistono punti di equilibrio, cioè caselle completamente segnate

La Tabella 12.5 riporta invece un esempio di gioco con equilibri multipli: sia la cella (1, 1) sia la cella (3, 2) sono equilibri di Nash.

	1	2	3
1	($\bar{15}$, $\bar{11}$)	(4, 1)	($\bar{6}$, 5)
2	(10, $\bar{4}$)	(7, 2)	(3, 1)
3	(3, 6)	($\bar{12}$, $\bar{8}$)	(5, 3)

Tabella 12.5: Tabella dei payoff con segnature per un gioco con equilibri multipli nelle posizioni (1, 1) e (3, 2)

12.3 La strategia del caso pessimo

In che modo un giocatore può scegliere la sua strategia quando il metodo dell'induzione a ritroso non fornisce una soluzione univoca? Il problema è analogo a quello della programmazione in condizioni di ignoranza, con la differenza che ciascun giocatore si trova di fronte non la natura, ma un altro giocatore razionale. Questo tuttavia non rende comunque ovvio quale scelta aspettarsi dall'avversario, dato che questi a sua volta potrà ragionare e adattare la sua scelta a quella del primo decisore, e viceversa, all'infinito.

Un approccio semplificativo tratto dalla programmazione in condizioni di ignoranza è il criterio del caso pessimo: il giocatore determina per ogni propria strategia la strategia dell'avversario che lo danneggia di più. Quindi, sceglie la strategia che, sotto questa ipotesi, gli garantisce il *payoff* massimo. Questo approccio in generale non fornisce la prestazione migliore, ma solo la prestazione più garantita: in generale, il risultato del gioco sarà migliore di quello previsto, dato che forse l'avversario non sceglierà proprio la strategia più dannosa per il decisore.

Definizione 50 Si definisce valore del gioco per il giocatore d il payoff massimo ottenibile dal giocatore d nel proprio caso pessimo, cioè la miglior garanzia possibile sulla sua prestazione:

$$u^{(d)} = \max_{x^{(d)} \in X^{(d)}} \min_{x^{(j)} \in X^{(j)}, \forall j \neq d} f^{(d)}(x^{(1)}, \dots, x^{(|D|)})$$

In dettaglio, la definizione valuta per ogni strategia $x^{(d)}$ del giocatore d il risultato per lui pessimo rispetto alle strategie adottate dagli altri giocatori, dunque il valore minimo del payoff $f^{(d)}$ rispetto alle strategie $x^{(j)}$ di tutti i giocatori $j \neq d$.

Approfondiremo in particolare il caso dei giochi con $|D| = 2$ giocatori. Come già detto, il giocatore $d = 1$ viene indicato come giocatore di riga e il giocatore $d = 2$ come giocatore di colonna: useremo gli indici (r) e (c) anziché (1) e (2) . Il valore del gioco per i due giocatori è definito come:

$$u^{(r)} = \max_{x^{(r)} \in X^{(r)}} \min_{x^{(c)} \in X^{(c)}} f^{(r)}(x^{(r)}, x^{(c)})$$

$$u^{(c)} = \max_{x^{(c)} \in X^{(c)}} \min_{x^{(r)} \in X^{(r)}} f^{(c)}(x^{(r)}, x^{(c)})$$

In generale, questi due valori sono del tutto scorrelati, dato che derivano da funzioni diverse e indipendenti. Vedremo però che per alcuni giochi essi sono invece correlati.

Esempio 79 Applichiamo la strategia del caso pessimo all'esempio della Tabella 12.3. I risultati sono riportati nella Tabella 12.6: per ogni riga si valuta l'impatto minimo rispetto alla funzione $f^{(r)}$, per ogni colonna l'impatto minimo rispetto alla funzione $f^{(c)}$.

$(f^{(r)}, f^{(c)})$	1	2	3	$\min f^{(r)}$
1	(4,3)	(5,1)	(6,2)	4
2	(2,1)	(8,4)	(3,3)	2
3	(5,9)	(9,6)	(2,8)	2

$\min f^{(c)}$	1	1	2

Tabella 12.6: Applicazione del criterio del caso pessimo

Quindi $u^{(r)} = 4$ e $u^{(c)} = 2$: giocando con il criterio del caso pessimo, il giocatore di riga vince almeno 4, qualunque sia la strategia seguita dal giocatore di colonna; giocando con il criterio del caso pessimo, il giocatore di colonna vince almeno 2, qualunque sia la strategia seguita dal giocatore di riga. In particolare, se entrambi seguono tale criterio, il primo guadagna 6 e il secondo 2. In generale, i guadagni potrebbero coincidere con i valori del gioco oppure essere superiori. In questo caso, uno dei due giocatori guadagna il minimo garantito, l'altro un valore strettamente più alto.

Il profilo $x_{\text{worst}} = (1, 3)$ generato dal criterio del caso pessimo non è di equilibrio: se il gioco fosse ripetuto, ciascuno dei due giocatori avrebbe un incentivo a scegliere una diversa strategia. Ovviamente, facendolo, perderebbe la garanzia sul caso pessimo. Se tale profilo fosse stato di equilibrio, i due giocatori non avrebbero avuto incentivi a modificare la propria scelta.

12.4 Giochi di interesse particolare

Due aspetti fondamentali nei giochi sono il valore complessivo (somma) dell'utilità distribuita ai giocatori e il modo in cui è distribuita. Con riferimento a tali aspetti, due categorie di giochi sono di particolare importanza:

1. *giochi a somma zero*: l'utilità complessiva del gioco è sempre nulla:

$$\sum_{d \in D} f^{(d)}(x^{(1)}, \dots, x^{(|D|)}) = 0 \text{ per ogni } (x^{(1)}, \dots, x^{(|D|)}) \in X$$

Si noti che cambiare la scala di misura delle utilità del gioco (unità di misura e posizione dello zero), cioè applicare una trasformazione $f' = af + b$ con $a > 0$ e b qualsiasi, non modifica le dominanze, né gli equilibri del gioco. Sono quindi da considerare giochi a somma zero anche tutti i giochi a somma uniforme $\sum_{d \in D} f'(x) = b|D|$.

2. *giochi simmetrici*: l'utilità di giocare una data strategia dipende solo dalle altre strategie giocate e non dal giocatore che le usa; quindi, permutando le strategie fra i giocatori, si permutano i *payoff* allo stesso modo:

$$f^{(d)}(x^{(1)}, \dots, x^{(|D|)}) = f^{(p_d)}(x^{(p_1)}, \dots, x^{(p_{|D|})}) \text{ per ogni } d \in D$$

e per ogni permutazione $p = (p_1, \dots, p_{|D|})$ degli indici $1, \dots, |D|$.

Capitolo 13

Giochi a somma zero

Un aspetto fondamentale nei giochi è il valore complessivo dell'utilità distribuita ai giocatori (cioè la somma dei *payoff* sui vari giocatori) e il modo in cui è distribuita nei vari profili di strategie. Un caso di particolare importanza è quello dei giochi a somma zero.

Definizione 51 Si definisce *gioco a somma zero* un gioco nel quale l'utilità complessiva è nulla in ogni profilo di strategie:

$$\sum_{d \in D} f^{(d)}(x^{(1)}, \dots, x^{(|D|)}) = 0 \text{ per ogni } (x^{(1)}, \dots, x^{(|D|)}) \in X$$

Si noti che cambiare la scala di misura delle utilità del gioco (unità di misura e posizione dello zero), cioè applicare una trasformazione $f' = af + b$ con $a > 0$ e b qualsiasi, non modifica le dominanze, né gli equilibri del gioco. Sono quindi da considerare giochi a somma zero anche tutti i giochi a somma uniforme $\sum_{d \in D} f'(x) = b|D|$.

Nei giochi a somma zero a due giocatori, *la vittoria di uno uguaglia la perdita dell'altro*. Di conseguenza, la matrice dei *payoff* del giocatore di colonna è opposta a quella del giocatore di riga, ed è ridondante indicarla. La Tabella 13.1 considera il caso di un gioco a due giocatori in cui entrambi hanno due strategie: a sinistra compare la rappresentazione completa, con i *payoff* di entrambi i giocatori; a destra quella compatta, con i *payoff* del solo giocatore di riga. Il gioco ha solo 4 valori indipendenti di guadagno, anziché 8.

	1	2
1	(a,-a)	(c,-c)
2	(b,-b)	(d,-d)

	1	2
1	a	c
2	b	d

Tabella 13.1: Un gioco a somma zero a due persone, rappresentato a sinistra in maniera completa, a destra in maniera compatta

Vediamo ora brevemente come cambiano i concetti fondamentali esposti nel capitolo precedente relativamente alla forma strategica, se si considerano i giochi a somma zero a due giocatori.

13.1 Strategie dominate

Per il giocatore di riga, la dominanza è definita come sempre: una strategia ne domina un'altra quando la riga corrispondente ha valori più alti, elemento per elemento. Per il giocatore di colonna, invece, i valori in tabella rappresentano costi, per cui una strategia ne domina un'altra quando la colonna corrispondente ha valori più bassi, elemento per elemento.

Esempio 80 La Figura 13.1 riporta tre esempi di gioco. Quello con la matrice F_1 non ha strategie di riga dominate, mentre la prima colonna domina la seconda (dunque, si può cancellare la seconda colonna e, a posteriori, concludere che il giocatore di riga sceglierà la seconda strategia). Nella matrice F_2 , per il giocatore di riga la terza strategia domina le altre due in quanto caratterizzata da guadagni maggiori; per il giocatore di colonna la terza strategia domina le altre due in quanto caratterizzata da perdite inferiori. Infine, nella matrice F_3 la prima riga è dominata, mentre non vi sono colonne dominate; eliminata la prima riga, la terza colonna risulta dominata dalla seconda; eliminata anche la terza colonna, la matrice risultante non ha dominanze.

$$F_1 = \begin{bmatrix} -2 & 3 \\ -1 & 1 \end{bmatrix} \quad F_2 = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 0 & -1 \\ 3 & 1 & 1 \end{bmatrix} \quad F_3 = \begin{bmatrix} -1 & 0 & -1 \\ 1 & 2 & 3 \\ 2 & -1 & -1 \end{bmatrix}$$

Figura 13.1: Tre esempi di giochi a somma zero a due persone

13.2 Equilibri

Definizione 52 In un gioco a somma zero a due persone, $(x^{*(r)}, x^{*(c)})$ è un punto di equilibrio se e solo se la casella (r, c) è un punto di sella della matrice dei payoff: punto di massimo per la colonna c e di minimo per la riga r :

$$\begin{cases} f_{rc} \geq f_{ic} \text{ per ogni } i \in X^{(r)} \\ f_{rc} \leq f_{rj} \text{ per ogni } j \in X^{(c)} \end{cases}$$

Esempio 81 Riprendendo i tre esempi della Figura 13.1, la matrice F_1 ammette il punto di sella $(2, 1)$ con $f_{21} = -1$. La matrice F_2 ha quattro punti di sella, che sono tutti gli elementi di valore 1. Infine, la matrice F_3 non ha punti di sella.

13.3 Valore del gioco

Nel caso dei giochi a somma zero a due persone, l'assunzione del caso pessimo è più ragionevole che altrove, perché l'impatto sui due giocatori è sempre esattamente opposto, e quindi il caso pessimo per un giocatore coincide col caso ottimo per l'altro. È quindi naturale ipotizzare che l'avversario giocherà in modo da produrre il danno massimo, e che il criterio del caso pessimo sia una guida utile.

Il concetto di valore del gioco per ciascuno dei due giocatori, cioè di miglior garanzia possibile, si definisce come prima per il giocatore di riga:

$$u^{(r)} = \max_{i \in X^{(r)}} \min_{j \in X^{(c)}} f_{ij}$$

mentre per il giocatore di colonna la definizione sarebbe:

$$\max_{i \in X^{(r)}} \min_{j \in X^{(c)}} (-f_{ij}) = \max_{i \in X^{(r)}} \left(-\max_{j \in X^{(c)}} f_{ij} \right) = -\min_{i \in X^{(r)}} \max_{j \in X^{(c)}} f_{ij}$$

Per convenzione, nei giochi a somma zero, si cambia il segno nella definizione del valore del gioco per il giocatore di colonna:

$$u^{(c)} = \min_{i \in X^{(r)}} \max_{j \in X^{(c)}} f_{ij}$$

Con questa convenzione, il valore del gioco descrive non un guadagno minimo, ma una perdita massima: il giocatore di colonna non massimizza il guadagno, ma minimizza la perdita.

Mentre in generale il valore del gioco per un giocatore non ha relazione con quello per l'altro, nei giochi a somma zero i due valori sono strettamente legati dal seguente teorema, di cui discuteremo poi una generalizzazione ancora più interessante.

Teorema 24 *Dato un gioco a somma zero a due persone, e i relativi valori del gioco per il giocatore di riga ($u^{(r)}$) e per quello di colonna ($u^{(c)}$):*

1. $u^{(r)} \leq u^{(c)}$;
2. $u^{(r)} = u^{(c)}$ se e solo se il gioco ha equilibri di Nash, cioè la matrice F ha un punto di sella.

Il teorema dice che il guadagno minimo garantito per il giocatore di riga non può superare la perdita massima garantita per il giocatore di colonna. Il motivo è abbastanza intuitivo, dato che si tratta di due limiti, uno inferiore e uno superiore, sulla stessa somma (incognita) che passa da una mano all'altra. I due limiti in genere non coincidono perché ciascuno dei due giocatori potrebbe avere un risultato migliore di quello garantito: un guadagno più alto, ovvero una perdita più bassa. Se F non ha punti di sella, il gioco non ammette equilibri: in ogni situazione, almeno uno dei giocatori ha un incentivo a cambiare strategia, dato che potrebbe migliorare il proprio risultato (nell'ipotesi che l'avversario non cambi strategia). D'altra parte, se e solo se F ha un punto di sella, le scelte prodotte dal criterio del caso pessimo sono coerenti e portano all'equilibrio che corrisponde al punto di sella. In tal caso, i due giocatori hanno valore del gioco identico, e questo viene detto *valore del gioco* senza ulteriori specificazioni. Esso è la massima utilità ottenibile da entrambi. Questo significa anche che, se vi sono più equilibri, ovvero punti di sella, tutti hanno lo stesso valore.

$$F_1 = \begin{bmatrix} -2 & 3 \\ -1 & 1 \end{bmatrix} \quad F_2 = \begin{bmatrix} -2 & -3 \\ 0 & 3 \end{bmatrix} \quad F_3 = \begin{bmatrix} 2 & -3 \\ 0 & 2 \\ -5 & 10 \end{bmatrix}$$

Figura 13.2: Tre esempi di giochi a somma zero a due persone

Esempio 82 La Figura 13.2 riporta altri tre esempi di giochi a somma zero a due persone. Per il primo gioco, il caso pessimo per il giocatore di riga corrisponde al payoff minimo su ciascuna riga, cioè -2 e -1 , e la strategia migliore è la seconda, da cui $u^{(r)}(F_1) = -1$. Per il giocatore di colonna, il caso pessimo corrisponde alla perdita massima su ciascuna colonna, cioè -1 e 3 , e la strategia migliore è la prima, da cui $u^{(c)}(F_1) = -1$.

Passando al secondo gioco, il caso pessimo per il giocatore di riga corrisponde al payoff minimo su ciascuna riga, cioè -3 e 0, e la strategia migliore è la seconda, da cui $u^{(r)}(F_2) = 0$. Per il giocatore di colonna, il caso pessimo corrisponde alla perdita massima su ciascuna colonna, cioè 0 e 3, e la strategia migliore è la prima, da cui $u^{(c)}(F_2) = 0$.

Infine, nel terzo gioco il caso pessimo per il giocatore di riga corrisponde al payoff minimo su ciascuna riga, cioè -3, 0 e -5, e la strategia migliore è la seconda, da cui $u^{(r)}(F_3) = 0$. Per il giocatore di colonna, il caso pessimo corrisponde alla perdita massima su ciascuna colonna, cioè 2 e 10, e la strategia migliore è la prima, da cui $u^{(c)}(F_3) = 2$.

Nelle prime due matrici i profili di strategie risultanti dal criterio del caso pessimo sono equilibri, mentre nella terza non vi sono equilibri e qualsiasi profilo di strategie dato fornisce ad almeno uno dei due giocatori un incentivo a cambiar strategia.

13.4 Strategie miste

Se il gioco viene giocato più volte in sequenza, diventa lecito per i giocatori cambiare strategia da una volta all'altra. In questa situazione, la variabile di decisione di ogni giocatore non avrà più un insieme finito di valori, ma un insieme continuo, che corrisponde alle frequenze con la quale il giocatore sceglie ciascuna alternativa fondamentale. Queste combinazioni si definiscono *strategie miste*, il che spiega finalmente perché le alternative fondamentali del gioco vengono chiamate strategie pure.

Questo modello descrive anche la situazione in cui lo stesso gioco viene giocato in parallelo da due squadre di giocatori, in cui i giocatori di ogni squadra possono adottare strategie diverse tra loro. E ancora, lo stesso modello descrive la situazione in cui i giocatori sono solo due, giocano un solo turno, e ognuno sceglie la propria alternativa con un'estrazione casuale, della quale ha prefissato le probabilità delle singole alternative possibili. In tutti questi casi, l'utilità dei giocatori diventa una variabile aleatoria, e lo scopo del gioco per il giocatore di riga è massimizzare il guadagno atteso, per quello di colonna è minimizzare la perdita attesa.

Definizione 53 Definiamo strategia mista per un giocatore un vettore di probabilità

$$\xi = [\xi, \dots, \xi_n]' \in \Xi = \left\{ \xi \in \mathbb{R}^n : \sum_{i=1}^n \xi_i = 1 \text{ e } \xi_i \geq 0 \text{ per ogni } i = 1, \dots, n \right\}$$

dove n indica il numero delle strategie pure disponibili per il giocatore.

In un gioco a due giocatori, quello di riga ha un insieme $X^{(r)}$ di $n^{(r)} = |X^{(r)}|$ possibili strategie pure, e quello di colonna un insieme $X^{(c)}$ di $n^{(c)} = |X^{(c)}|$ possibili strategie pure. Combinandole con opportune probabilità, possono ottenere le strategie miste $\xi^{(r)} \in \Xi^{(r)}$ e $\xi^{(c)} \in \Xi^{(c)}$. Determinare l'impatto atteso di queste strategie per i due giocatori è facile: il guadagno atteso per il primo coincide con la perdita attesa per il secondo e si ottiene sommando per ogni combinazione di esiti $(i, j) \in \Xi^{(r)} \times \Xi^{(c)}$ il prodotto del guadagno f_{ij} per la probabilità $\xi_i^{(r)} \xi_j^{(c)}$ associati alla combinazione.

Definizione 54 Dato un gioco F a due persone a somma zero e due strategie $\xi^{(r)}$ e $\xi^{(c)}$, definiamo valore atteso del gioco il valore atteso del guadagno per il giocatore

di riga e della perdita per il giocatore di colonna.

$$v(F) = E \left[f \left(\xi^{(r)}, \xi^{(c)} \right) \right] = \sum_{i=1}^{n^{(r)}} \sum_{j=1}^{n^{(c)}} \xi_i^{(r)} \xi_j^{(c)} f_{ij}$$

Se i due giocatori adottano il criterio del caso pessimo, quello di riga massimizza il minimo del valore atteso, mentre quello di colonna minimizza il massimo del valore atteso. Entrambi, però, operano sui vettori di probabilità $\xi^{(r)}$ e $\xi^{(c)}$, anziché sulla semplice scelta di una strategia pura.

Definizione 55 Definiamo valore atteso del giocatore di riga *il valore atteso ottimo del suo guadagno nel caso pessimo*:

$$v^{(r)}(F) = \max_{\xi^{(r)} \in \Xi^{(r)}} \min_{\xi^{(c)} \in \Xi^{(c)}} E \left[f \left(\xi^{(r)}, \xi^{(c)} \right) \right]$$

e valore atteso del giocatore di colonna *il valore atteso ottimo della sua perdita nel caso pessimo*:

$$v^{(c)}(F) = \min_{\xi^{(c)} \in \Xi^{(c)}} \max_{\xi^{(r)} \in \Xi^{(r)}} E \left[f \left(\xi^{(r)}, \xi^{(c)} \right) \right]$$

Si noti che questi valori non coincidono con i valori $u^{(r)}$ e $u^{(c)}$ introdotti nella Definizione 50 della Sezione 12.3, dato che quelli si riferiscono a strategie pure, mentre questi a strategie miste. Tecnicamente parlando, i problemi che definiscono questi valori attesi non sono altro che rilassamenti continui dei corrispondenti problemi che definiscono i valori del gioco rispetto alle strategie pure. Infatti, una strategia pura è un caso particolare di strategia mista in cui una probabilità è unitaria e le altre sono nulle. La funzione obiettivo è la stessa, i vincoli di somma unitaria e di non negatività sono identici, ma una strategia pura rispetta in più un vincolo di integralità. La relazione fra u e v non è ovvia, dato che nel secondo caso entrambi i giocatori hanno più possibilità di scelta: da un lato, ogni giocatore ha maggiori possibilità di migliorare il proprio risultato, ma dall'altro le possibilità di essere danneggiati dal proprio avversario sono molto maggiori. Per semplificare questo studio, è possibile fare un'osservazione importante.

Proposizione 4 Qualunque strategia mista un giocatore adotti, il caso pessimo per lui sarà rappresentato da una strategia pura dell'avversario. Di conseguenza, il valore atteso del gioco per il giocatore di riga si può esprimere come:

$$v^{(r)}(F) = \max_{\xi^{(r)} \in \Xi^{(r)}} \min_{j \in X^{(c)}} E \left[f_{\cdot j}(\xi^{(r)}) \right] \text{ con } E \left[f_{\cdot j}(\xi^{(r)}) \right] = \sum_{i=1}^{n^{(r)}} \xi_i^{(r)} f_{ij}$$

e il valore atteso del gioco per il giocatore di colonna come:

$$v^{(c)}(F) = \min_{\xi^{(c)} \in \Xi^{(c)}} \max_{i \in X^{(r)}} E \left[f_{i \cdot}(\xi^{(c)}) \right] \text{ con } E \left[f_{i \cdot}(\xi^{(c)}) \right] = \sum_{j=1}^{n^{(c)}} \xi_j^{(c)} f_{ij}$$

Dimostrazione. Consideriamo il caso del giocatore di riga. Fissata la strategia $\xi^{(r)}$, il suo *payoff* nel caso pessimo sarà

$$\phi(\xi^{(r)}) = \min_{\xi^{(c)} \in \Xi^{(c)}} \sum_{i=1}^{n^{(r)}} \sum_{j=1}^{n^{(c)}} \xi_i^{(r)} \xi_j^{(c)} f_{ij} = \min_{\xi^{(c)} \in \Xi^{(c)}} \sum_{j=1}^{n^{(c)}} \left(\sum_{i=1}^{n^{(r)}} \xi_i^{(r)} f_{ij} \right) \xi_j^{(c)}$$

dove il termine fra parentesi tonde è un vettore di coefficienti costanti con indice j , dato che le probabilità $\xi_i^{(r)}$ che definiscono la strategia mista del giocatore di riga sono fissate. Si noti anche che ciascuno di tali coefficienti è il valore atteso della corrispondente colonna della matrice dei payoff:

$$\sum_{i=1}^{n^{(r)}} \xi_i^{(r)} f_{ij} = E [f_{\cdot j}(\xi^{(r)})]$$

cioè il valore atteso del payoff ottenuto dal giocatore di riga contro ciascuna delle strategie pure del giocatore di colonna.

Ora determiniamo il caso pessimo rispetto alla strategia $\xi^{(r)}$ che abbiamo fissato. Per farlo, si deve risolvere il problema di ottimizzazione:

$$\begin{aligned} \phi(\xi^{(r)}) &= \min_{\xi^{(c)} \in \Xi^{(c)}} \sum_{j=1}^{n^{(c)}} E [f_{\cdot j}(\xi^{(r)})] \xi_j^{(c)} \\ &\quad \sum_{j=1}^{n^{(c)}} \xi_j^{(c)} = 1 \\ &\quad \xi_j^{(c)} \geq 0 \end{aligned}$$

Si tratta di minimizzare una combinazione convessa di $n^{(c)}$ numeri. La soluzione ottima di tale problema è banalmente il minimo di tali numeri:

$$\xi_j^{(c)}(\xi^{(r)}) = \begin{cases} 1 & \text{per } j = \arg \min_{j \in X^{(c)}} E [f_{\cdot j}(\xi^{(r)})] \\ 0 & \text{altrimenti} \end{cases}$$

cioè consiste nel determinare l'impatto pessimo e fissare a 1 la probabilità della strategia avversaria che lo produce. Il valore corrispondente è

$$\phi(\xi^{(r)}) = \min_{j \in X^{(c)}} E [f_{\cdot j}(\xi^{(r)})]$$

A questo punto, il valore del gioco per il giocatore di riga si ottiene fissando la strategia mista che massimizza $\phi(\xi^{(r)})$, da cui la tesi. Con lo stesso procedimento si ottiene il valore del gioco per il giocatore di colonna. ■

Questo teorema mostra che, ponendosi nell'ottica prudente del caso pessimo, si deve considerare l'avversario come se si comportasse in modo deterministico, scegliendo ogni volta la strategia pura più dannosa per la strategia mista scelta da noi, anziché in modo stocastico scegliendo una combinazione di strategie. Infatti, se una strategia pura è più dannosa delle altre, anche se l'avversario potrebbe "annacquarla" combinandola con loro, non ha senso che lo faccia dal punto di vista del caso pessimo.

13.5 Teorema del minimax di Von Neumann

Teorema 25 *Per ogni gioco finito a somma zero a due persone, i valori attesi del gioco per il giocatore di riga ($v^{(r)}$) e per il giocatore di colonna ($v^{(c)}$) coincidono. Inoltre, esiste una coppia di strategie miste $(\xi^{*(r)}, \xi^{*(c)}) \in \Xi^{(r)} \times \Xi^{(c)}$ tali che*

$$u^{(r)} = u^{(c)} = E [f(\xi^{*(r)}, \xi^{*(c)})]$$

Queste due strategie sono ottimali per i due giocatori, cioè

$$\begin{aligned} u^{(r)} &= \min_{\xi^{(c)} \in \Xi^{(c)}} E \left[f(\xi^{*(r)}, \xi^{(c)}) \right] \\ u^{(c)} &= \max_{\xi^{(r)} \in \Xi^{(r)}} E \left[f(\xi^{(r)}, \xi^{*(c)}) \right] \end{aligned}$$

Dimostrazione. Dimostreremo questo teorema riconoscendo la relazione che lega la teoria dei giochi a somma zero e la teoria della dualità nella Programmazione Lineare. Consideriamo per primo il punto di vista del giocatore di riga. Questi vuole scegliere il vettore di probabilità $\xi^{(r)}$ che definisce la sua strategia mista, in modo da massimizzare la propria utilità nel caso pessimo. Da un punto di vista matematico, questo problema si può rappresentare come un problema di massimizzazione nelle variabili di decisione $\xi_i^{(r)}$ ($i = 1, \dots, n^{(r)}$):

$$\begin{aligned} \max \phi(\xi^{(r)}) \\ \sum_{i=1}^{n^{(r)}} \xi_i^{(r)} = 1 \\ \xi_i^{(r)} \geq 0 \quad i = 1, \dots, m \end{aligned}$$

$$\text{con } \phi(\xi^{(r)}) = \min_{j \in X^{(c)}} \sum_{i=1}^{n^{(r)}} f_{ij} \xi_i^{(r)}.$$

Questo problema ha vincoli lineari sulle variabili $\xi_i^{(r)}$, ma una funzione obiettivo non lineare, dato che contiene l'operatore di minimizzazione (min). Tuttavia, è possibile trasformare il problema in un problema lineare equivalente con un semplice trucco. Si aggiunge una nuova variabile v che descrive l'utilità nel caso pessimo. Poi si impone che v non superi nessuno dei termini $\sum_{i=1}^{n^{(r)}} f_{ij} \xi_i^{(r)}$. In questo modo, v potrebbe essere più piccola di tutti quei termini, ma poiché il problema tende a massimizzare v , in qualsiasi soluzione ottima sicuramente varrà un'uguaglianza, cioè v coinciderà con il termine minimo.

In questo modo si ottiene un problema equivalente di Programmazione Lineare (sia i vincoli sia la funzione obiettivo sono lineari):

$$\begin{aligned} \max_{v, \xi^{(r)}} \phi(v, \xi^{(r)}) &= v \\ \sum_{i=1}^{n^{(r)}} \xi_i^{(r)} &= 1 \\ v - \sum_{i=1}^{n^{(r)}} f_{ij} \xi_i^{(r)} &\leq 0 \quad j = 1, \dots, n^{(c)} \\ \xi_i^{(r)} &\geq 0 \quad i = 1, \dots, n^{(r)} \end{aligned}$$

Per il giocatore di colonna, il problema è analogo, salvo che bisogna minimizzare una perdita. Si applicherà un trucco molto simile per linearizzare la funzione

obiettivo, rimuovendo un operatore di massimizzazione (max):

$$\begin{aligned} \min_{w, \xi^{(c)}} \psi(w, \xi^{(c)}) &= w \\ \sum_{j=1}^{n^{(c)}} \xi_j^{(c)} &= 1 \\ w - \sum_{j=1}^{n^{(c)}} f_{ij} \xi_j^{(c)} &\geq 0 \quad i = 1, \dots, n^{(r)} \\ \xi_j^{(c)} &\geq 0 \quad j = 1, \dots, n^{(c)} \end{aligned}$$

I due problemi sono duali, in quanto:

1. uno è di massimizzazione e l'altro di minimizzazione;
2. ciascuno ha come vettore dei coefficienti dell'obiettivo il vettore dei termini noti dei vincoli dell'altro e come vettore dei termini noti dei vincoli quello dei coefficienti dell'obiettivo dell'altro: sono entrambi vettori costituiti da un 1 accompagnato da una serie di 0;
3. ciascuno ha matrice di coefficienti uguale alla matrice dei coefficienti dell'altro trasposta: la prima riga è costituita da uno 0 e una serie di 1, le successive da un 1 e dai coefficienti f_{ij} cambiati di segno; poiché le sommatorie sono una sull'indice i e l'altra sull'indice j , i vincoli di uno dei due problemi scorrono colonne della matrice F e i vincoli dell'altro le righe.

La teoria della dualità nella Programmazione Lineare garantisce che i due problemi hanno entrambi soluzione ottima, e che i valori dei due ottimi sono identici, cioè il massimo che il giocatore di riga può ottenere è uguale al minimo che il giocatore di colonna può perdere. Questo è esattamente l'enunciato della tesi da dimostrare. ■

Il teorema del minimax di Von Neumann ha tre conseguenze principali:

1. le strategie miste offrono a entrambi i giocatori garanzie migliori rispetto alle strategie pure¹. Banalmente, siccome le strategie pure sono casi particolari di strategie miste nelle quali una ha probabilità unitaria e le altre probabilità nulla, le strategie miste non possono fare peggio.
2. le strategie miste ottimali hanno lo stesso valore per i due giocatori, e danno luogo a un equilibrio (misto), che è migliore dei valori del gioco per le strategie pure:

$$u^{(r)} \leq v^{(r)} = E \left[f(\xi^{\circ(r)}, \xi^{\circ(c)}) \right] = v^{(c)} \leq u^{(c)}$$

3. se il gioco ha un punto di sella f_{rc} , allora le strategie miste ottimali coincidono con le strategie pure nelle quali si giocano la riga e la colonna associate alla sella.

$$\exists \text{ punto di sella } x^\circ = (x^{\circ(r)}, y^{\circ(r)}) \Rightarrow \begin{cases} \xi_i^{\circ(r)} \in \{0, 1\} \text{ per } i \in X^{(r)} \\ \xi_j^{\circ(c)} \in \{0, 1\} \text{ per } j \in X^{(c)} \\ u^{(r)} = v^{(r)} = v^{(c)} = u^{(c)} \end{cases}$$

¹Vero è che la garanzia offerta dalle strategie pure è deterministica, mentre quella offerta dalle strategie miste riguarda il valore atteso.

13.5.1 Determinazione della strategia mista di equilibrio

Se il gioco è particolarmente semplice, cioè ciascuno dei due giocatori ha solo due strategie, la ricerca delle strategie miste di equilibrio richiede di risolvere dei problemi di ottimizzazione con due variabili (le probabilità $\xi_1^{(r)}$ e $\xi_2^{(r)}$ per il giocatore di riga e le probabilità $\xi_1^{(c)}$ e $\xi_2^{(c)}$ per quello di colonna). Siccome la somma di tali probabilità è unitaria, in effetti si tratta di semplici problemi monodimensionali, che si possono risolvere anche graficamente. Vediamo nel seguito alcuni esempi.

Esempio 83 Il gioco del pari e dispari è un gioco a somma zero a due giocatori, ciascuno dei quali ha due strategie. La matrice dei payoff, riportata in Tabella 13.2, non contiene equilibri.

	P	D
P	1	-1
D	-1	1

Tabella 13.2: Matrice dei payoff del gioco del pari e dispari

Sappiamo che il caso pessimo per qualsiasi strategia è determinato da una strategia pura dell'avversario. Attribuiamo al giocatore di riga una generica strategia mista $\xi^{(r)} = [\alpha \ (1 - \alpha)]^T$, e valutiamo i risultati delle due possibili strategie pure dell'avversario:

- $E [f(\xi^{(r)}, P)]$ vale $\alpha \cdot 1 + (1 - \alpha) \cdot (-1) = 2\alpha - 1$;
- $E [f(\xi^{(r)}, D)]$ vale $\alpha \cdot (-1) + (1 - \alpha) \cdot 1 = 1 - 2\alpha$.

Il problema del giocatore di riga è massimizzare il peggiore dei due risultati

$$v^{(r)} = \max_{\alpha \in [0;1]} \min (2\alpha - 1, 1 - 2\alpha)$$

La parte sinistra della Figura 13.3 illustra la risoluzione grafica del problema per il giocatore di riga: per ogni sua strategia mista, cioè per ogni valore di α , si considerano le due strategie pure del giocatore di colonna, rappresentate da due rette. Le strategie miste per il giocatore di colonna ricadrebbero nella regione compresa fra le due rette, mentre la funzione $\phi^{(r)}(\xi^{(r)})$, che rappresenta il caso pessimo, corrisponde all'inviluppo inferiore di tale regione, evidenziato in linea continua. A questo punto, il giocatore di riga deve decidere la propria strategia mista ottimale $\xi^{(r)}$, che è data dal massimo della funzione lineare a tratti $\phi^{(r)}(\xi^{(r)})$. Quindi, la strategia ottima è $\xi^{(r)} = [1/2 \ 1/2]^T$, e il valore del gioco è nullo. In effetti, la strategia migliore per il gioco è giocare metà delle volte pari e metà dispari; giocare più spesso una delle due strategie espone al rischio di essere scoperti e prevenuti. Il guadagno medio di tale strategia ottimale è nullo.

Il problema del giocatore di colonna si risolve considerando la generica strategia mista $\xi^{(c)} = [\beta \ (1 - \beta)]^T$ e valutandone il caso pessimo. Ancora una volta, ci si può limitare a considerare le due strategie pure dell'avversario:

- $E [f(P, \xi^{(c)})]$ vale $\beta \cdot 1 + (1 - \beta) \cdot (-1) = 2\beta - 1$;
- $E [f(D, \xi^{(c)})]$ vale $\beta \cdot (-1) + (1 - \beta) \cdot 1 = 1 - 2\beta$;

dato che quelle miste hanno risultati intermedi. Quindi, il giocatore di colonna deve minimizzare il peggiore dei due risultati

$$v^{(c)} = \min_{\beta \in [0;1]} \max (2\beta - 1, 1 - 2\beta)$$

La parte destra della Figura 13.3 illustra la risoluzione grafica del problema: la funzione obiettivo è rappresentata dall'inviluppo superiore delle due rette. Il punto di massimo è individuato dall'intersezione delle due rette, che corrisponde alla strategia ottima $\xi^{*(c)} = [1/2 \ 1/2]^T$, per la quale il valore del gioco è nullo.

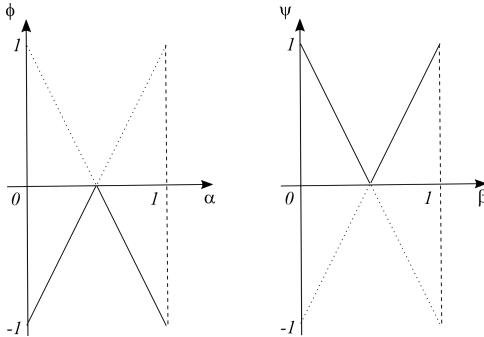


Tabella 13.3: Risoluzione grafica del gioco del pari e dispari per il giocatore di riga (a sinistra) e di colonna (a destra)

Esempio 84 Consideriamo ora la matrice di payoff riportata in Tabella 13.4.

	1	2
1	2	-3
2	-1	1

Tabella 13.4: Matrice dei payoff per un gioco a somma zero

Le strategie miste per il giocatore di riga si possono rappresentare come $\xi^{(r)} = [\alpha \ 1 - \alpha]^T$ con $\alpha \in [0, 1]$. Per ogni valore di α , il caso pessimo corrisponde alla peggiore delle due strategie pure disponibili per il giocatore di colonna, rappresentate nella parte sinistra della Figura 13.3 dalle due rette:

- $E[f(\xi^{(r)}, 1)] = \alpha \cdot 2 + (1 - \alpha) \cdot (-1) = 3\alpha - 1$;
- $E[f(\xi^{(r)}, 2)] = \alpha \cdot (-3) + (1 - \alpha) \cdot 1 = 1 - 4\alpha$.

A questo punto, per ogni strategia mista $\xi^{(r)}$, cioè per ogni α , il caso pessimo è rappresentato dall'inviluppo inferiore della regione compresa fra le due rette, evidenziato in linea continua. Il giocatore di riga deve decidere la strategia mista che massimizza questo risultato:

$$v^{(r)} = \max_{\alpha \in [0, 1]} \min (3\alpha - 1, 1 - 4\alpha)$$

Tale strategia è individuata dall'intersezione delle due rette, per cui $\xi^{\circ(r)} = [2/7 \ 5/7]^T$ e il valore del gioco è $v^{(r)} = -1/7$. Con lo stesso ragionamento, dalla parte destra della Figura 13.3 si ricava che $\xi^{\circ(c)} = [4/7 \ 3/7]^T$ e $v^{(c)} = -1/7$.

13.5.2 Metodo del pivot*

C'è un altro modo per trasformare il problema in un programma lineare che rende più semplice i calcoli nel caso in cui il valore del gioco sia positivo. D'altra parte,

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

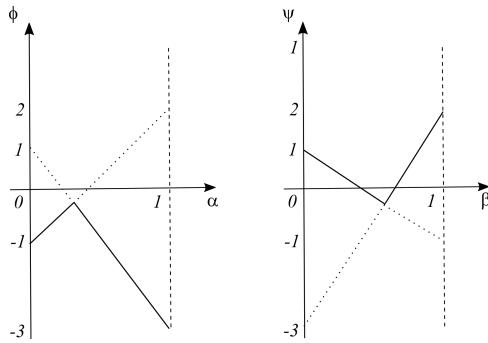


Figura 13.3: Risoluzione grafica del gioco di Tabella 13.4 per il giocatore di riga (a sinistra) e di colonna (a destra)

se si aggiunge una costante a tutti i payoff della matrice, il gioco sostanzialmente non cambia, e si può certamente rendere positivo il suo valore. Ipotizziamo quindi che sia $v > 0$ e riformuliamo il problema del giocatore di riga:

$$\begin{aligned} \max_{v, \xi^{(r)}} \phi(v, \xi^{(r)}) &= v \\ \sum_{i=1}^{n^{(r)}} \xi_i^{(r)} &= 1 \\ v - \sum_{i=1}^{n^{(r)}} f_{ij} \xi_i^{(r)} &\leq 0 \quad j = 1, \dots, n^{(c)} \\ \xi_i^{(r)} &\geq 0 \quad i = 1, \dots, n^{(r)} \end{aligned}$$

Introduciamo le variabili ausiliarie $z_i = \xi_i^{(r)}/v$, da cui $\xi_i^{(r)} = vz_i$.

$$\begin{aligned} \max_{v, z} \phi(v, z) &= v \\ \sum_{i=1}^{n^{(r)}} vz_i &= 1 \\ v - \sum_{i=1}^{n^{(r)}} f_{ij} vz_i &\leq 0 \quad j = 1, \dots, n^{(c)} \\ z_i &\geq 0 \quad i = 1, \dots, n^{(r)} \end{aligned}$$

Il vincolo $\sum_{i=1}^{n^{(r)}} z_i = 1/v$ non è lineare, ma possiamo sfruttare il fatto che massimizzare v equivale a minimizzare $1/v$, per rimuovere completamente v dal problema minimizzando al suo posto $\sum_{i=1}^{n^{(r)}} z_i$. Negli altri vincoli, infatti, v compare come fattore moltiplicativo a entrambi i membri, e quindi si può semplificare:

$$\begin{aligned} \max_z \phi(v, z) &= \sum_{i=1}^{n^{(r)}} z_i \\ \sum_{i=1}^{n^{(r)}} f_{ij} z_i &\geq 1 \quad j = 1, \dots, n^{(c)} \\ z_i &\geq 0 \quad i = 1, \dots, n^{(r)} \end{aligned}$$

Risolto questo problema, la soluzione del gioco originale può essere facilmente trovata tenendo conto che $v = 1/\sum_{i=1}^{n(r)} z_i^\circ$ e $\xi_i^{(r)} = v z_i^\circ$. Per il giocatore di colonna, si può applicare un procedimento del tutto analogo.

Per risolvere il problema, si può applicare un metodo equivalente all'algoritmo del simplex²

Step 1 Se necessario, aggiungere una costante a tutti gli elementi della matrice di utilità affinché il valore del gioco sia positivo (a questo punto, il valore del gioco originale è uguale a quello del nuovo gioco meno la costante aggiunta).

0	-1	-1	...	-1	
1	a_{11}	a_{12}	...	a_{1n}	x_1
1	a_{21}	a_{22}	...	a_{2n}	x_2
:	:	:	:	:	:
1	a_{m1}	a_{m2}	...	a_{mn}	x_m
	y_1	y_2	...	y_n	

Step 2 Creare un *tableau* aumentando la matrice delle utilità (coefficienti delle variabili x nei vincoli) con una riga di -1 in alto (coefficienti della funzione obiettivo cambiati di segno) e una colonna di 1 all'estrema sinistra (termini noti dei vincoli). Nell'angolo in alto a sinistra porre 0 (termine noto della funzione obiettivo). Aggiungere una colonna a destra con le etichette da x_1 a x_m (variabili fuori base) e una riga in basso con le etichette da y_1 a y_n (variabili basiche). Sia $A = [a_{ij}]$ la nuova tabella.

Step 3 Selezionare un elemento a_{pq} , detto *pivot*, tale che

1. $q \in \{j : a_{0,j} < 0, 1 \leq j \leq n\}$
2. $p = \arg \min_{1 \leq i \leq m, a_{i,q} > 0} \left\{ \frac{a_{i,0}}{a_{i,q}} \right\}$.

Nota: il pivot esiste sempre perchè il problema è limitato sotto le condizioni espresse dal teorema minmax.

Step 4 Fare pivot nel seguente modo:

1. $a_{ij} := a_{ij} - a_{pj} \cdot a_{iq}/a_{pq} \quad i = 1, \dots, m \quad j = 1, \dots, n \quad i \neq p \quad j \neq q$
2. nella riga del pivot, ad eccezione del pivot, $a_{pj} := \frac{a_{pj}}{a_{pq}} \quad j = 1, \dots, n \quad j \neq q$
3. nella colonna del pivot, ad eccezione del pivot, $a_{iq} := -\frac{a_{iq}}{a_{pq}} \quad i = 1, \dots, m \quad i \neq p$
4. $a_{pq} := \frac{1}{a_{pq}}$.

Step 5 Scambiare l'etichetta a destra nella riga del *pivot* (la variabile esce di base) con l'etichetta in basso nella colonna del *pivot* (la variabile entra in base), swap $A_{p,n+1} \ A_{m+1,q}$.

²La parte seguente non è stata ancora rivista.

Step 6 Se rimane qualche elemento negativo nella riga zero, tornare allo *step 3*.

Step 7 Altrimenti, ricavare la soluzione nel seguente modo:

1. il valore del gioco v è il reciproco del numero nell'angolo in alto a sinistra (valore della funzione obiettivo all'ottimo)
2. la strategia ottima del giocatore riga è così costruita: le variabili x_i che si trovano nella colonna a destra ricevono probabilità 0 (variabili fuori base), mentre quelle sulla riga in basso una probabilità pari a $v \cdot a_{0,i}$ (variabili basiche)
3. la strategia ottima del giocatore colonna è così costruita (soluzioni duali del problema appena risolto): le variabili y_j che si trovano nella riga in basso ricevono probabilità 0, mentre quelle sulla colonna a destra una probabilità pari a $v \cdot a_{j,0}$.

Esempio 85 Consideriamo il gioco descritto dalla matrice

$$B = \begin{bmatrix} 2 & -1 & 6 \\ 0 & 1 & -1 \\ -2 & 2 & 1 \end{bmatrix}$$

Come si può notare non esiste un punto di sella e quindi nessuna strategia pura ottimale. Il valore del gioco è positivo? Forse. In ogni caso è più semplice sommare 2 ad ogni elemento e ottenere la matrice.

$$B' = \begin{bmatrix} 4 & 1 & 8 \\ 2 & 3 & 1 \\ 0 & 4 & 3 \end{bmatrix}$$

Il valore di questo gioco è almeno 1 dal momento che il giocatore riga si può garantire questa vittoria scegliendo la riga 1 o 2. Questo completa lo step 1.

Nello step 2, prepariamo un tableau per B' come il seguente:

0	-1	-1	-1	
1	4	1	8	x_1
1	2	3	1	x_2
1	0	4	3	x_3
	y_1	y_2	y_3	

Nello step 3, dobbiamo scegliere il pivot. Poiché tutte e tre le colonne hanno nella prima riga un valore negativo, possiamo sceglierne una qualunque come colonna del pivot. Prendiamo la prima. La riga del pivot deve avere un numero positivo su questa colonna, così deve essere una delle prime due righe. Per decidere quale è, calcoliamo il rapporto tra il numero nella prima colonna e il candidato pivot. Per la prima riga è $1/4$, per la seconda $1/2$. Il più piccolo è il primo quindi scegliamo 4 come pivot.

0	-1	-1	-1	
1	4	1	8	x_1
1	2	3	1	x_2
1	0	4	3	x_3
	y_1	y_2	y_3	

Nello step 4, dopo aver combinato linearmente le righe del tableau, otteniamo

$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{3}{4}$	1	
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	2	x_1
$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{5}{2}$	-3	x_2
1	0	4	3	x_3
	y_1	y_2	y_3	

Nello step 5, scambiamo le etichette della riga e della colonna del pivot, cioè x_1 e y_1 .

$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{3}{4}$	1	
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	2	y_1
$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{5}{2}$	-3	x_2
1	0	4	3	x_3
	x_1	y_2	y_3	

Nello step 6, controlliamo che ci siano valori negativi nella prima riga. Poichè ci sono torniamo allo step 3.

Rifacendo tutti i passaggi precedenti otteniamo

0.4	0.1	0.3	0.1	
0.2	0.3	-0.1	2.3	y_1
0.2	-0.2	0.4	-1.2	y_2
0.2	0.8	-1.6	7.8	x_3
	x_1	x_2	y_3	

Nello step 7, possiamo leggere la soluzione del gioco con la matrice B' :

- il valore del gioco è 2.5.
- la strategia ottima per il giocatore riga è $\mathbf{p} = (0.25, 0.75, 0)$
- la strategia ottima per il giocatore colonna è $\mathbf{q} = (0.5, 0.5, 0)$

Il gioco con matrice B ha le stesse strategie ottime, ma valore $2.5 - 2 = 0.5$.

Capitolo 14

Giochi simmetrici

Un altro caso di particolare importanza nella distribuzione dell'utilità fra i giocatori è quello in cui i giocatori sono perfettamente intercambiabili, cioè hanno le stesse strategie a disposizione e, se si scambiano le strategie, ne ricavano gli stessi *payoff*.

**** SONO ARRIVATO QUI ****

Definizione 56 Definiamo gioco simmetrico un gioco nel quale tutti i giocatori hanno lo stesso numero di strategie pure $n = |X^{(d)}|$ (per ogni $d \in D$) e, per ogni permutazione $p = (p_1, \dots, p_{|D|})$ dei giocatori, se si applica p alle strategie adottate, i payoff risultanti sono anch'essi permutati in base a p :

$$f^{(d)}(x^{(1)}, \dots, x^{(|D|)}) = f^{(p_d)}(x^{(p_1)}, \dots, x^{(p_{|D|})}) \text{ per ogni } d \in D$$

Intuitivamente, questo significa che i risultati delle strategie non dipendono da chi le gioca: scambiando le strategie fra i giocatori, si scambiano allo stesso modo i risultati. Da un punto di vista pratico, tipicamente questo corrisponde al fatto che ogni giocatore ha a disposizione strategie che in qualche modo corrispondono *una a una* a quelle di ogni altro giocatore, e che quindi è conveniente nominare e ordinare allo stesso modo. Quindi, la prima strategia di ciascun giocatore corrisponderà in qualche senso alla prima strategia di ciascun altro giocatore, la seconda alla seconda, e così via. Non sono necessariamente le stesse operazioni fisiche, ma sono operazioni che producono gli stessi risultati.

Nel seguito consideriamo il caso dei giochi simmetrici a due persone. In questo caso, esistono solo due possibili permutazioni, e quindi la definizione si semplifica molto:

$$f_{ij}^{(r)} = f_{ji}^{(c)} \text{ per ogni } i, j = 1, \dots, n$$

Se il giocatore di riga ottiene il guadagno $f_{ij}^{(r)}$ applicando la strategia i mentre quello di colonna applica la strategia j , quando i due giocatori si scambiano le strategie e il giocatore di riga applica j mentre quello di colonna applica i , sarà il giocatore di colonna a ottenere lo stesso guadagno.

Osservazione 9 In un gioco simmetrico a due persone, la matrice dei guadagni del secondo giocatore è la trasposta della matrice dei guadagni del primo giocatore: $F_2 = F'_1$.

Contrariamente al caso dei giochi a somma zero, anche se l'informazione associata al secondo giocatore è ridondante, in genere si usa comunque la rappresentazione completa. Questo perché l'informazione mancante non sarebbe semplicemente l'opposto di quella riportata, ma andrebbe cercata nella casella simmetrica, con un

procedimento molto meno evidente. La Tabella 14.1 riporta un esempio di gioco simmetrico a due persone.

	1	2	3
1	(2,2)	(3,0)	(5,2)
2	(0,3)	(4,4)	(-1,8)
3	(2,5)	(8,-1)	(6,6)

Tabella 14.1: Esempio di gioco simmetrico a due persone: le matrici dei payoff dei due giocatori sono una la trasposta dell'altra

14.1 Tassonomia dei giochi simmetrici a due persone e due strategie

Un caso estremamente semplificato, ma comunque molto interessante in pratica, è quello dei giochi a due persone in cui ogni giocatore ha solo due strategie a disposizione. È possibile classificare tali giochi in base a due diverse metodologie, una delle quali porta a 4 classi, mentre l'altra porta a 12 classi. Le due metodologie sono largamente coerenti, nel senso che la seconda aggiunge una determinazione ulteriore alla prima, spezzando alcune delle sue 4 classi in sottoclassi. Nel seguito, presenteremo le due classificazioni in termini astratti, e poi discuteremo alcuni giochi cui la letteratura attribuisce suggestivi nomi simbolici, spiegando le loro caratteristiche fondamentali nei termini delle due classificazioni.

14.1.1 Classificazione in base agli equilibri di Nash

	1	2
1	(f_{11}, f_{11})	(f_{12}, f_{21})
2	(f_{21}, f_{12})	(f_{22}, f_{22})

Tabella 14.2: Generico gioco simmetrico a due persone e due strategie: le matrici dei payoff dei due giocatori sono una la trasposta dell'altra

Data la generica matrice dei *payoff* in Tabella 14.2, ed escludendo i casi limite nei quali alcuni *payoff* coincidono, sono possibili solo quattro situazioni. Per ciascuna valutiamo se ammette strategie dominate e determiniamo gli eventuali equilibri:

1. $f_{11} > f_{21} > f_{12} > f_{22}$: la prima strategia domina strettamente la seconda per entrambi i giocatori; esiste un solo equilibrio di Nash nel punto $(1, 1)$;
2. $f_{11} > f_{21} \text{ e } f_{12} < f_{22}$: non vi sono strategie dominanti, e vi sono due equilibri di Nash giacenti sulla diagonale principale, nei punti $(1, 1)$ e $(2, 2)$;
3. $f_{11} < f_{21} \text{ e } f_{12} > f_{22}$: non vi sono strategie dominanti, e vi sono due equilibri di Nash giacenti sulla diagonale secondaria, nei punti $(1, 2)$ e $(2, 1)$;
4. $f_{11} < f_{21} \text{ e } f_{12} < f_{22}$: la seconda strategia domina strettamente la prima per entrambi i giocatori; esiste un solo equilibrio di Nash nel punto $(2, 2)$.

Come si può notare, *i giochi simmetrici a due persone e due strategie hanno sempre almeno un equilibrio*. Questa è una caratteristica notevole, dato che in generale

Classe	Sottoclasse	Ordinamento	Esempi
1	a	$f_{11} > f_{12} > f_{21} > f_{22}$	Matrimonio perfetto
	b	$f_{11} > f_{12} > f_{22} > f_{21}$	
	c	$f_{11} > f_{21} > f_{12} > f_{22}$	
	d	$f_{12} > f_{11} > f_{21} > f_{22}$	
	e	$f_{12} > f_{11} > f_{22} > f_{21}$	
2	a	$f_{11} > f_{21} > f_{22} > f_{12}$	Caccia al cervo
	b	$f_{11} > f_{22} > f_{12} > f_{21}$	Coordinamento (1)
	c	$f_{11} > f_{22} > f_{21} > f_{12}$	Coordinamento (2)
3	a	$f_{21} > f_{11} > f_{12} > f_{22}$	Corsa del coniglio
	b	$f_{12} > f_{21} > f_{11} > f_{22}$	Guerra dei sessi (1)
	c	$f_{21} > f_{12} > f_{11} > f_{22}$	Guerra dei sessi (2)
4	a	$f_{21} > f_{11} > f_{22} > f_{12}$	Dilemma del prigioniero

Tabella 14.3: Elenco dei 12 possibili giochi simmetrici a due giocatori con due strategie secondo la classificazione per ordinamento dei payoff e loro relazione con la classificazione secondo gli equilibri di Nash

ciò non è vero né per i giochi asimmetrici né per quelli simmetrici con più di due strategie (per esempio, il gioco della morra cinese). Vedremo in seguito, considerando alcuni esempi notevoli, che significato pratico assumano questi equilibri.

14.1.2 Classificazione in base all'ordinamento dei payoff

Un modo più dettagliato di classificare i giochi simmetrici a due persone e due strategie si basa sull'ordinamento relativo dei quattro *payoff* indipendenti, f_{11} , f_{12} , f_{21} e f_{22} . In questa classificazione, vi sono tanti giochi quante sono le permutazioni di quattro valori, cioè $4! = 24$. In realtà, però, è possibile ridurre queste classi a 12 osservando che i nomi delle strategie 1 e 2 sono del tutto convenzionali, per cui non si lede la generalità della classificazione facendo l'ipotesi aggiuntiva $f_{11} > f_{22}$. Se ci troviamo di fronte a un gioco che non soddisfa questa condizione, basta scambiare i nomi delle due strategie per renderla valida.

La Tabella 14.1.2 elenca i 12 possibili giochi secondo questa tassonomia, specificando l'ordinamento dei *payoff* e l'appartenenza di ogni classe con le quattro classi della prima tassonomia, nonché alcuni nomi celebri di giochi descritti in letteratura che ricadono in ciascuna sottoclasse.

Le sezioni seguenti presentano sei giochi notevoli, scelti fra le 12 categorie. Per rendere la descrizione più concreta, invece di usare *payoff* simbolici, attribuiremo ai guadagni i valori convenzionali 0, 1, 2 e 3. Per lo stesso motivo, attribuiremo alle due strategie i nomi convenzionali di *cooperazione* (*C*) per la prima e *non cooperazione* (*NC*) per la seconda. Questo è coerente con l'ipotesi che due strategie cooperanti abbiano un *payoff* f_{11} superiore al *payoff* f_{22} di due strategie non cooperanti.

14.2 Il matrimonio perfetto

Questo gioco corrisponde all'ordinamento

$$f_{11} > f_{21} > f_{12} > f_{22}$$

ed è rappresentato nella Tabella 14.4.

	C	NC
C	$(\bar{3}, \bar{3})$	$(\bar{1}, 2)$
NC	$(2, \bar{1})$	$(0, 0)$

Tabella 14.4: Matrice dei *payoff* per il matrimonio perfetto: le marcature sui *payoff* migliori per ogni riga e colonna indicano che il punto (C, C) è l'unico equilibrio di Nash

Questo gioco descrive quindi la situazione in cui la cooperazione aumenta la soddisfazione dei due giocatori, come in un matrimonio ideale. Entrando nei dettagli, l'ordinamento fra le possibili situazioni è:

1. se entrambi i coniugi/giocatori cooperano (C, C) entrambi guadagnano il miglior risultato possibile;
2. se il primo giocatore non coopera mentre l'altro coopera (NC, C) , guadagna il secondo miglior risultato possibile;
3. se il primo giocatore coopera mentre l'altro non coopera (C, NC) , guadagna il terzo risultato;
4. se entrambi non cooperano (NC, NC) , entrambi guadagnano il risultato peggiore.

La strategia di non cooperare (NC) è dominata dalla strategia di cooperare (C) per entrambi i giocatori. Questo significa che è ragionevole per entrambi cooperare, e infatti il gioco ha un solo equilibrio di Nash, che corrisponde alla mutua cooperazione (nella Tabella 14.4 entrambi i payoff sono marcati).

Della classe 1 (giochi con un solo equilibrio cooperativo) fanno parte altri quattro giochi, abbastanza simili, che non hanno ispirato nomi immaginosi alla letteratura, probabilmente perché corrispondono a situazioni, in cui cooperare con una controparte che non coopera è più conveniente che non cooperare con una controparte che coopera ($f_{12} > f_{21}$). Pare poco probabile che queste situazioni si presentino spontaneamente, senza un intervento esterno.

14.3 La caccia al cervo

Questo gioco corrisponde all'ordinamento

$$f_{11} > f_{21} > f_{22} > f_{12}$$

ed è rappresentato nella Tabella 14.5.

	C	NC
C	$(\bar{3}, \bar{3})$	$(0, 2)$
NC	$(2, 0)$	$(\bar{1}, \bar{1})$

Tabella 14.5: Tabella dei payoff per la caccia al cervo: le marcature sui payoff migliori per ogni riga e colonna indicano che i punti (C, C) e (NC, NC) sono entrambi equilibri di Nash

Questo gioco deriva da un passo del *Discorso sull'origine delle disuguaglianze tra gli uomini* (1755) di Rousseau¹, secondo il quale le società umane derivano dalle alleanze temporanee fra uomini per la caccia a grandi animali, sui quali un individuo isolato non potrebbe avere la meglio. Supponendo che alla caccia partecipino due giocatori, l'ordinamento fra le possibili situazioni è:

1. se entrambi cooperano (C, C), è molto probabile che arrivino alla cattura di un cervo, che è il miglior risultato ottenibile per entrambi;
2. se uno dei due non coopera, mentre l'altro coopera (NC, C), quello che non coopera può dedicare parte del suo tempo a catturare una preda minore, (una lepre) e mantenere una piccola speranza di catturare anche il cervo con lo sforzo residuo, ottenendo quindi il secondo miglior risultato possibile;
3. se entrambi i giocatori non cooperano (NC, NC), entrambi possono catturare una lepre, ma la probabilità che poi riescano a prendere il cervo è praticamente nulla, per cui questo è il terzo risultato possibile per entrambi;
4. infine, se uno dei due giocatori coopera, mentre l'altro no (C, NC), quello che coopera guadagna solo una minima speranza di prendere il cervo, che è il risultato peggiore possibile.

Come si vede nella Tabella 14.5, nessuna delle due strategie è dominata, e il gioco ha due equilibri di Nash, che corrispondono alla mutua cooperazione e alla mutua non cooperazione. Infatti, nel caso in cui si possa essere certi che l'altro giocatore cooperi, conviene non abbandonare l'equilibrio cooperativo, perché porta alla cattura del cervo. Invece, nel caso si possa essere certi che l'altro giocatore non cooperi, conviene non abbandonare l'equilibrio non cooperativo e mantenere almeno la cattura di una lepre. Ci sono dunque due possibilità simmetriche. La mutua cooperazione è un equilibrio desiderabile perché corrisponde a guadagni più alti per entrambi. La mutua non cooperazione è però anch'essa un equilibrio, benché meno desiderabile per entrambi, ed è conservativa, cioè obbedisce al criterio del caso pessimo. Quindi, entrambe le strategie sono razionali, e si sceglierà l'una o l'altra strategia in base ad altri fattori non modellati nella forma base del gioco, come il livello di fiducia verso l'altro giocatore.

14.4 I giochi di coordinamento puro

Questi giochi corrispondono agli ordinamenti

$$f_{11} > f_{22} > f_{12} > f_{21} \quad \text{e} \quad f_{11} > f_{22} > f_{21} > f_{12}$$

e sono rappresentati nella Tabella 14.6. Ovviamente, il caso intermedio in cui $f_{12} = f_{21}$ ricade nella stessa categoria, ed è anzi la rappresentazione più comune di questo genere di gioco. Questi giochi descrivono le situazioni nelle quali i risultati migliori si ottengono quando entrambi i giocatori scelgono la stessa strategia, mentre le strategie asimmetriche sono dannose per entrambi. In questo senso, le etichette *cooperare* e *non cooperare* sono poco adeguate, e infatti non vengono usate nella tabella.

Un esempio classico è quello di due automobilisti che si incontrano, venendo da direzioni opposte, su una strada stretta, senza segnaletica orizzontale. Se entrambi si tengono sul lato destro, oppure sul lato sinistro, passano senza problemi. Si può pensare che una delle due soluzioni sia leggermente migliore dell'altra (per esempio,

¹Jean-Jacques Rousseau (1712-1778), filosofo, scrittore e musicista svizzero.

	1	2		1	2
1	$(\bar{3}, \bar{3})$	$(0, 1)$		$(\bar{3}, \bar{3})$	$(1, 0)$
2	$(1, 0)$	$(\bar{2}, \bar{2})$		$(0, 1)$	$(\bar{2}, \bar{2})$

Tabella 14.6: Tabella dei payoff per i giochi di puro coordinamento: le marcature sui payoff migliori per ogni riga e colonna indicano che i punti $(1, 1)$ e $(2, 2)$ sono entrambi equilibri di Nash

se i due automobilisti sono abituati a guidare sulla destra, sarà leggermente più comodo che entrambi si tengano a destra). Invece, se uno dei due automobilisti si tiene sulla destra e l’altro sulla sinistra, il risultato è uno scontro frontale, cioè un risultato cattivo per entrambi. Anche in questo caso, si può pensare che uno dei due possa avere un risultato leggermente migliore dell’altro, ma comunque peggiore che nel caso di strategie coordinate.

I giochi di coordinamento puro non hanno strategie dominate e hanno gli stessi due punti di equilibrio della caccia al cervo. La differenza sta nel fatto che la contrapposizione fra l’equilibrio più redditizio e il meno redditizio è meno spiccata che nella caccia al cervo: sono situazioni meno interessanti.

14.5 La corsa del coniglio

Questo gioco corrisponde all’ordinamento

$$f_{21} > f_{11} > f_{12} > f_{22}$$

ed è rappresentato nella Tabella 14.7.

	C	NC
C	$(2, 2)$	$(\bar{1}, \bar{3})$
NC	$(\bar{3}, \bar{1})$	$(0, 0)$

Tabella 14.7: Tabella dei payoff per la corsa del coniglio: le marcature sui payoff migliori per ogni riga e colonna indicano che i punti (NC, C) e (C, NC) sono entrambi equilibri di Nash

Questo gioco deriva dal film *Gioventù bruciata* (1955): due giovani si sfidano guidando ognuno un’auto rubata a tutta velocità verso un precipizio; quello che salta fuori per primo dall’automobile è un “coniglio”, cioè un vigliacco (nell’originale inglese, si parla di *chicken*), e perde la sfida. Un nome alternativo per lo stesso gioco è *falchi e colombe*, espressione coniata da Russell² nel suo libro *Il senso comune e la guerra nucleare* (1959), per descrivere la situazione di tensione politica tra USA e URSS. In queste sfide (militari, politiche, economiche, ecc...), uno dei due contendenti gioca il tutto per tutto nella speranza che l’altro ceda, rischiando il comune disastro.

L’ordinamento fra le possibili situazioni è:

- se un giocatore non coopera e l’altro sì (NC, C) , cioè il primo insiste e il secondo si ritira, quello che non coopera vince la sfida, e ottiene il miglior risultato possibile;

²Bertrand Arthur William Russell (1872-1970), filosofo e matematico galleso.

- se entrambi i giocatori cooperano (C, C), cioè si ritirano, entrambi guadagnano il secondo risultato possibile, cioè un pareggio onorevole;
- se un giocatore coopera, cioè si ritira, mentre l'altro non coopera (C, NC), quello che coopera ottiene il terzo possibile risultato, cioè viene svergognato, ma salva la vita;
- se entrambi i giocatori non cooperano (NC, NC), cioè insistono, entrambi perdono la vita, che è il peggior risultato possibile per entrambi.

Come si vede nella Tabella 14.7, nessuna delle due strategie è dominata, e il gioco ha due equilibri di Nash, che corrispondono a strategie complementari, con la cooperazione di uno e la non cooperazione dell'altro giocatore. Infatti, nel caso in cui un giocatore potesse essere certo che l'altro coopererà, gli converrebbe non cooperare. Viceversa, se potesse essere certo che l'altro non coopererà, gli converrebbe cooperare. Il criterio del caso pessimo suggerisce a entrambi i giocatori di cooperare, ma la soluzione risultante non è di equilibrio, e quindi tende a non conservarsi nel caso di ripetizioni del gioco. Nonostante esistano ben due equilibri di Nash non esiste alcuna strategia razionale per il gioco.

14.6 La guerra dei sessi

Questo gioco corrisponde agli ordinamenti

$$f_{12} > f_{21} > f_{11} > f_{22} \quad \text{e} \quad f_{21} > f_{12} > f_{11} > f_{22}$$

ed è rappresentato nella Tabella 14.8.

	1	2		1	2
1	(1, 1)	$(\bar{3}, \bar{2})$	1	(1, 1)	$(\bar{2}, \bar{3})$
2	$(\bar{2}, \bar{3})$	(0, 0)	2	$(\bar{3}, \bar{2})$	(0, 0)

Tabella 14.8: Tabella dei payoff per la guerra dei sessi: le marcature sui payoff migliori per ogni riga e colonna indicano che i punti $(1, 2)$ e $(2, 1)$ sono entrambi equilibri di Nash

Come nei giochi di coordinamento, le etichette “cooperazione” e “non cooperazione” sono poco adeguate, e infatti non vengono utilizzate. Il gioco può essere esemplificato come segue. Due innamorati hanno deciso di trascorrere la giornata insieme. Discutono animatamente al telefono se andare a vedere la partita di calcio, come vorrebbe lui, o il balletto classico, come vorrebbe lei. La comunicazione si interrompe senza possibilità di ristabilirla, per cui ognuno deve decidere che cosa fare indipendentemente. Entrambi preferiscono uscire insieme piuttosto che rimanere separati, dato che la solitudine renderebbe sgradevole anche lo spettacolo preferito.

L'ordinamento delle situazioni possibili è il seguente:

- il miglior risultato si ottiene stando insieme a guardare il proprio spettacolo preferito, quindi nel caso $(1, 2)$ per il primo giocatore e nel caso $(2, 1)$ per il secondo;
- il secondo miglior risultato si ottiene nel caso $(2, 1)$, stando insieme a guardare lo spettacolo preferito dall'altro, quindi nel caso $(2, 1)$ per il primo giocatore e nel caso $(1, 2)$ per il secondo;

- il terzo possibile risultato si ottiene nel caso $(1, 1)$, andare a vedere ciascuno il proprio spettacolo preferito;
- il risultato peggiore, infine, si ottiene nel caso $(2, 2)$, andando a vedere da soli lo spettacolo meno gradito.

Questo gioco non ha strategie dominate, e ha due punti di equilibrio, corrispondenti alle strategie asimmetriche, come nella corsa del coniglio. La differenza sta nel fatto che tali equilibri sono fondamentalmente positivi per entrambi i giocatori, invece di essere positivi per uno e molto negativi per l'altro.

14.7 Il dilemma del prigioniero

Questo gioco corrisponde all'ordinamento

$$f_{21} > f_{11} > f_{22} > f_{12}$$

ed è rappresentato nella Tabella 14.9.

	C	NC
C	$(2, 2)$	$(0, \bar{3})$
NC	$(\bar{3}, 0)$	$(\bar{1}, \bar{1})$

Tabella 14.9: Tabella dei payoff per il dilemma del prigioniero: le marcature sui payoff migliori per ogni riga e colonna indicano che il punto (NC, NC) è l'unico equilibrio di Nash

Questo gioco deriva da un lavoro di Flood³ e Dresher⁴ scritto nel 1950 per la *Rand Corporation* (*Applicazioni ad una strategia nucleare globale*). Il nome del gioco si deve però a Tucker⁵, insieme al seguente racconto per illustrarlo: due ladri, arrestati e imprigionati in due celle separate, sono sospettati anche di un efferato delitto. Il giudice ha prove sufficienti per dimostrare che i due sono colpevoli del reato minore, ma per infliggere una pena più severa avrebbe bisogno della confessione del reato maggiore. Il giudice offre quindi ai due ladri, mantenuti rigorosamente isolati, uno sconto di pena a chi accettasse di testimoniare contro l'altro. In caso di accettazione, chi testimonia subirà una condanna molto lieve, mentre l'altro ladro andrà in carcere per il massimo della pena. Se entrambi confessano avranno solo la concessione di alcune attenuanti generiche. Se entrambi tacciono, scamperanno la condanna per il reato maggiore e saranno imprigionati solo per il reato minore.

L'ordinamento fra le possibili situazioni è:

- se un giocatore non coopera e l'altro sì (NC, C) , cioè il primo confessa e il secondo tace, quello che confessa ottiene la riduzione forte di pena, che è il miglior risultato possibile;
- se entrambi i giocatori cooperano (C, C) , cioè tacciono (la cooperazione è fra i giocatori, non con il giudice, che è un soggetto esterno), entrambi sono condannati a una pena lieve, che è il secondo miglior risultato possibile;

³Merrill Meeks Flood (1908-1991), matematico americano.

⁴Melvin Dresher (1911-1992), matematico americano nato in Polonia.

⁵Albert William Tucker (1905-1995), il matematico canadese che propose anche le condizioni di KKT.

- se entrambi i giocatori non cooperano fra loro (NC, NC), cioè confessano, entrambi vengono condannati con le attenuanti generiche, che è il terzo risultato possibile per entrambi;
- se un giocatore coopera e l'altro no (C, NC), quello che coopera, cioè tace, riceve la condanna maggiore, che è il risultato peggiore possibile.

La strategia di cooperare (C) è dominata dalla strategia di non cooperare (NC) per entrambi i giocatori. Questo significa che è ragionevole per entrambi confessare, e infatti il gioco ha un solo equilibrio di Nash, che corrisponde alla mutua non cooperazione (nella tabella, entrambi i payoff sono marcati). Benché la soluzione sia dimostrata a filo di logica, ad essa si accompagna un dilemma. Infatti, la non cooperazione è la sola alternativa non dominata, e quindi razionale. Tuttavia, la mutua cooperazione sarebbe più desiderabile per entrambi i giocatori, perché $f_{22} > f_{11}$). Il dilemma formalizza un conflitto fra razionalità e interesse generale, che sembra essere tipico di molte situazioni concrete. Il dilemma del prigioniero è probabilmente in assoluto il gioco più studiato.

14.8 Altri giochi

Dei giochi generici (non simmetrici e non a somma zero) diremo poco. Ci limitiamo a descrivere un caso abbastanza curioso di gioco asimmetrico, che dà luogo a conseguenze paradossali. Tali conseguenze sono però sostenute non solo dalla teoria matematica, ma anche da osservazioni empiriche di situazioni che ricadono nel modello.

14.8.1 Il gioco dei maiali nel recinto

Questo gioco, detto anche gioco dei *piccioni in gabbia*, è un gioco non simmetrico e non a somma zero. La Tabella 14.10 riporta i payoff.

	Leva	Attesa
Leva	(4, 2)	(3, 3)
Attesa	(5, 0)	(1, 1)

Tabella 14.10: Tabella dei payoff per il gioco dei maiali nel recinto: le marcature sui payoff migliori per ogni riga e colonna indicano che il punto (Leva, Attesa) è l'unico equilibrio di Nash

Il gioco contempla un giocatore più forte (1) e uno più debole (2). Una leva consente di accedere a un premio (per esempio, del cibo). Il premio, però, è accessibile solo da una posizione lontana dalla leva (la parte opposta del recinto). Ognuno dei due giocatori può decidere se agire sulla leva o aspettare:

- se entrambi agiscono sulla leva, il giocatore forte raggiunge il cibo prima dell'altro e se ne accaparra la maggior parte;
- se il giocatore forte agisce sulla leva e quello debole aspetta vicino alla posizione dove compare il cibo, entrambi ne acquistano una buona porzione (quello forte scaccia il debole, ma questo ha il tempo di procurarsene una buona quantità per primo);

- se il giocatore debole agisce sulla leva e quello forte aspetta vicino alla posizione dove compare il cibo, il giocatore forte acquisisce quasi tutto il cibo, usando i vantaggi della posizione e della forza;
- se entrambi aspettano vicino alla posizione, il cibo non compare ed entrambi hanno il guadagno pessimo.

Il giocatore forte (di riga nella Tabella 14.10) non ha strategie dominate, mentre per il giocatore debole (di colonna) la strategia di agire sulla leva è dominata da quella di attendere. Quindi, il giocatore debole attende. A quel punto, per il giocatore forte è ragionevole agire sulla leva. La conclusione, paradossale, è che il giocatore forte ha convenienza ha mettersi al servizio del giocatore debole, dato che comunque riuscirà a procurarsi le briciole, mentre l'altro non ha speranza di riuscirvi se non attendendo.

14.9 Giochi finiti e strategie miste

Nash estese la teoria di Von Neumann e Morgenstern sulle strategie miste dai giochi a somma zero a tutti i giochi con insiemi finiti di giocatori e strategie. Brevemente, i risultati degni di nota sono due:

1. *ogni gioco finito ammette almeno un equilibrio in strategie miste;*
2. non è però possibile usare la Programmazione Lineare per trovare tali equilibri: sono noti algoritmi per determinare un equilibrio di Nash risolvendo una sequenza di problemi lineari la cui soluzione converge a un equilibrio, ma il numero degli equilibri può essere esponenziale rispetto al numero di giocatori e strategie e i diversi equilibri possono avere valori diversi (nei giochi a somma zero hanno tutti lo stesso valore, e spesso ce n'è uno solo); al momento per enumerare gli equilibri di Nash occorrono algoritmi di enumerazione implicita esponenziali.

14.10 Exercises* † ‡

Esercizio 1

È dato il gioco a due persone e a somma nulla rappresentato dalla seguente matrice di *payoff*, dove il giocatore di riga A punta a massimizzare e il giocatore di colonna B a minimizzare.

f	b_1	b_2
a_1	50	-20
a_2	-10	20
a_3	-50	10

Si formuli e si risolva il problema di determinare la strategia ottima (mista, se necessario) per il giocatore di riga A e la sua vincita attesa.

Si ricavi quindi la strategia ottima (anch'essa mista, se necessario) per il giocatore di colonna B .

Supponendo che il giocatore B segua una strategia meccanica che sceglie casualmente b_1 con probabilità 0.75 e b_2 con probabilità 0.25 e che il giocatore A lo sappia, qual è la vincita minima garantita per il giocatore A ? Motivare la risposta.

Soluzione

La strategia a_3 è dominata da a_2 , per cui non verrà mai usata. Volendo costruire una strategia mista ottimale che combini a_1 e a_2 occorre determinare la probabilità α di giocare a_1 in modo da massimizzare il risultato atteso nel caso pessimo:

$$\max [50\alpha - 10(1-\alpha), -20\alpha + 20(1-\alpha)] = \max [60\alpha - 10, -40\alpha + 20] \text{ con } \alpha \in [0; 1]$$

La risoluzione grafica suggerisce che $\alpha^* = 0.3$, da cui $f^* = 8$.

Corrispondentemente, il giocatore di colonna minimizza il risultato atteso nel caso per lui pessimo:

$$\min [50\beta - 20(1-\beta), -10\beta + 20(1-\beta)] = \min [70\beta - 20, -30\beta + 20] \text{ con } \beta \in [0; 1]$$

da cui $\beta^* = 0.4$, che conferma $f^* = 8$.

Se il giocatore B si comporta meccanicamente, è possibile considerarlo come uno stato di natura e applicare il criterio del valore atteso. Siccome con le probabilità date è $E[f(a_1)] = 32.5$, $E[f(a_2)] = -2.5$ e $E[f(a_3)] = -35$, conviene scegliere a_1 .

Esercizio 2

Dato il gioco finito a due persone a somma nulla rappresentato dalla seguente matrice dei benefici per il giocatore di riga, determinare la strategia mista ottimale per il giocatore di riga.

f	b_1	b_2	b_3	b_4
a_1	50	10	90	30
a_2	70	120	40	80

Determinare quindi la strategia mista ottimale per il giocatore di colonna.

*The solutions of these exercises have not yet been revised: error reports are welcome.

†I owe several of these exercises to exam texts of professor Alberto Colomni.

‡Diversi esercizi sono molto più complicati di quelli che verranno proposti all'esame o riguardano argomenti più generali di quelli svolti durante il corso. Ho deciso di tenerli per completezza, segnalandoli.

Soluzione

La strategia ottima per il giocatore di riga è $\xi_{a_1}^* = 0.6$ e $\xi_{a_2}^* = 0.4$, con $f^* = 60$. La strategia ottima per il giocatore di colonna è $\xi_{b_1}^* = x_{b_2}^* = 0$, e $\xi_{b_3}^* = x_{b_4}^* = 0.5$, sempre con $f^* = 60$.

Esercizio 3

Si consideri il gioco a due persone (A e B) a somma nulla definito dalle seguenti regole:

- si parte fissando $s = 0$;
- il giocatore A muove per primo, sommando 1 o 2 a s ;
- il giocatore B muove per secondo, sommando 1 o 2 a s ;
- i giocatori muovono a turni, prima A poi B , finché uno dei due rende $s \geq 5$: quel giocatore vince il gioco;
- se vince A , riceve 100 (e B perde 100); se vince B , riceve 300 (e A perde 300).

Si individuino le strategie possibili per ciascuno dei due giocatori.

Si descriva il gioco in forma estesa (come albero) e in forma strategica (come matrice), indicando il valore delle etichette per ciascun nodo dell'albero e per ciascuna casella della matrice (profilo di strategie).

In base alla descrizione in forma strategica, si dica se esistono strategie dominate e se esistono punti di equilibrio.

Soluzione

Poiché ad ogni passo s cresce almeno di 1 e il gioco si conclude quando $s \geq 5$, si termina al massimo alla quinta mossa (tre turni). La strategia di ogni giocatore è quindi descritta da una coppia di scelte (tutte le mosse nel terzo turno hanno lo stesso impatto). La matrice dei risultati è la seguente, e riporta sulle righe le strategie di A , sulle colonne le strategie di B e nelle caselle le vincite di A (le vincite di B sono esattamente opposte, dato che il gioco è a somma zero).

	(1, 1)	(1, 2)	(2, 1)	(2, 2)
(1, 1)	100	-300	-300	-300
(1, 2)	-300	100	100	100
(2, 1)	-300	-300	100	100
(2, 2)	100	100	100	100

La strategia $(2, 2)$ per il giocatore A domina tutte le altre e gli garantisce la vincita di 100. Per il giocatore B la strategia $(1, 2)$ domina le altre. Esiste quindi il punto di equilibrio $((2, 2), (1, 2))$.

Esercizio 4

Dato il gioco finito a due persone con la seguente matrice di *payoff*, si determinino le eventuali strategie dominate e gli eventuali equilibri di Nash.

f	C	D	E
A	(3, 2)	(5, 4)	(7, 8)
B	(5, 9)	(1, 11)	(4, 3)

Soluzione

La strategia C è dominata dalla strategia D . Eliminatala, la strategia B diventa a sua volta dominata. A questo punto, la strategia D diventa a sua volta dominata e rimane il solo profilo (A, E) , che è anche un equilibrio di Nash.

Esercizio 5

Si determinino le eventuali strategie dominate e gli eventuali equilibri di Nash del gioco finito a due persone con la seguente matrice di *payoff*.

f	B_1	B_2	B_3
A_1	(0,1)	(-1,3)	(1,3)
A_2	(0,2)	(1,1)	(2,0)
A_3	(1,0)	(1,-1)	(3,0)

Soluzione

La strategia A_3 domina le altre due strategie per il giocatore di riga. Il giocatore di colonna non ha inizialmente strategie dominate, ma dopo l'eliminazione di A_1 e A_2 , la strategia B_2 risulta dominata dalle altre. Rimangono i profili (A_3, B_1) e (A_3, B_3) , che sono anche equilibri di Nash.

Esercizio 6

Alle elezioni si presentano due schieramenti, i Buoni e i Cattivi. Ciascuno schieramento presenta un candidato in ogni circoscrizione. Ci sono tre circoscrizioni: Nord, Centro e Sud. Entrambi gli schieramenti hanno un leader, che possono candidare al massimo in due delle tre circoscrizioni. La tabella che segue fornisce una stima, compiuta in base ai sondaggi, dei milioni di voti che i candidati (il leader o un altro candidato) prenderebbero se si presentassero in ciascuna delle tre circoscrizioni.

f	Buoni		Cattivi	
	Leader	Altro	Leader	Altro
Nord	16	14	19	13
Centro	18	14	15	13
Sud	17	14	18	14

In ogni circoscrizione, il candidato con più voti ottiene per il suo schieramento 10 deputati per ogni milione di voti in più rispetto all'avversario. Gli strateghi dei due schieramenti devono decidere dove candidare i loro leader, rispettando il limite massimo di due circoscrizioni.

Si descriva il problema come un gioco finito a due persone a somma nulla, fornendo la matrice dei payoff per i Buoni.

Si dica se esistono scelte dominate e se esiste un punto di equilibrio o no.

Soluzione

Anzi tutto, è chiaro che conviene candidare il leader in due circoscrizioni, dato che attira sempre più voti dei candidati alternativi. Quindi entrambi i giocatori hanno esattamente tre strategie, ciascuna caratterizzata dalla circoscrizione in cui non si candida il leader. Ne deriva la seguente matrice di *payoff*.

f	Nord	Centro	Sud
Nord	30	-10	10
Centro	10	-30	-10
Sud	20	-20	0

Il profilo di strategie (Nord,Centro), cioè con i Buoni che candidano il loro leader al Centro e al Sud e i Cattivi che candidano il loro al Nord e al Sud, è di equilibrio, con vincita pari a 10 per i Cattivi. Le altre righe e colonne sono tutte dominate.

Esercizio 7

Il gioco dell'OX si svolge tra due giocatori nel seguente modo. Il giocatore A sceglie Pari (P) oppure Dispari (D) e comunica la sua scelta all'avversario; poi, il giocatore B sceglie un numero tra 2 e 5 e lo dichiara. Se A ha scelto P , paga all'avversario tante unità quante ne ha dichiarato B, se ha scelto D riceve la stessa quantità dall'avversario. Il gioco deve essere obbligatoriamente ripetuto in "set" di quattro partite, durante le quali ogni giocatore deve usare le sue scelte in modo equilibrato: il primo, cioè, deve scegliere due volte P e due volte D , mentre il secondo deve scegliere una volta ciascuno dei valori interi compresi fra 2 e 5.

Si descriva l'albero di gioco per una singola partita.

Si elenchino le strategie di entrambi i giocatori per una singola partita.

Si mostri un esempio di percorso sull'albero completo, associando alla foglia il risultato corrispondente.

Si indichino il numero di livelli e di foglie per un intero "set" di quattro partite, specificando quante sono le foglie ammissibili⁶.

Si valuti se il gioco offre ad entrambi i giocatori le stesse possibilità di vincita complessiva su un set di quattro partite⁷.

Soluzione

L'albero di gioco ha la seguente struttura.

-	P	5	(-5, 5)
		4	(-4, 4)
		3	(-3, 3)
		2	(-2, 2)
	D	5	(5, -5)
		4	(4, -4)
		3	(3, -3)
		2	(2, -2)

Nella singola partita, il primo giocatore ha due strategie (P e D), il secondo ne ha quattro per ogni mossa del primo, dunque 8. Un esempio di strategia per il secondo giocatore è giocare 5 contro P e 2 contro D .

Un esempio di cammino lungo l'intero albero è $(P, 5, P, 4, D, 3, D, 2)$ che produce un *payoff* di $(-4, 4)$.

Su un intero set, l'albero ha 8 livelli: quelli gestiti dal primo giocatore hanno due figli per nodo, quelli gestiti dal secondo hanno un numero di figli decrescente,

⁶Domanda un po' più sofisticata.

⁷Domanda più sofisticata.

da 4 a 1. Quindi, le foglie sono: $2 \cdot 4 \cdot 2 \cdot 3 \cdot 2 \cdot 2 \cdot 2 \cdot 1 = 2^4 \cdot 4! = 384$. Tuttavia, la regola del bilanciamento riduce le scelte del primo giocatore da 2^4 a $\binom{4}{2}$, per cui le foglie ammissibili sono solo 144.

Il gioco non è equilibrato, perché B può vincere sempre sulla singola partita, ma anche sulle quattro, scegliendo il valore massimo ancora disponibile in corrispondenza a P e il valore minimo in corrispondenza a D .

Esercizio 8⁸

Si determinino i punti di equilibrio (se ne esistono) del seguente gioco continuo a due persone a somma nulla:

$$\begin{cases} f_1(x_1, x_2) = \frac{1}{3}x_1^3 + x_1x_2^2 - 2x_1 - \frac{2}{3}x_2^3 & \text{con } x_1 \in X^{(1)} = \mathbb{R} \text{ e } x_2 \in X^{(2)} = \mathbb{R} \\ f_2(x_1, x_2) = -f_1(x_1, x_2) \end{cases}$$

Soluzione

Un punto di equilibrio, se esiste, è punto di massimo per il giocatore 1 e di minimo per il giocatore 2. Nel continuo, si possono determinare massimi e minimi con le note condizioni di annullamento sulla derivata prima:

$$\begin{cases} \frac{\partial f_1}{\partial x_1} = x_1^2 + x_2^2 - 2 = 0 \\ \frac{\partial f_1}{\partial x_2} = 2x_1x_2 - 2x_2^2 = 0 \end{cases}$$

e di segno sulla derivata seconda

$$\begin{cases} \frac{\partial^2 f_1}{\partial x_1^2} = 2x_1 < 0 \\ \frac{\partial^2 f_1}{\partial x_2^2} = 2x_1 - 4x_2 > 0 \end{cases}$$

Le soluzioni con $x_2 = 0$ sono inammissibili ($2x_1 < 0$ e $2x_1 > 0$). Quelle con $x_2 \neq 0$ hanno $x_2 = x_1$, e quindi $x_1 = x_2 = \pm 1$. Siccome $x_1 - 2x_2 > 0$, l'unico punto di equilibrio è $(-1, -1)$.

Esercizio 9

Si consideri il gioco continuo a due persone a somma nulla con funzione di beneficio $f(a, b) = -2a^2 + b^2 + 3ab - a - 2b$ per il giocatore A che sceglie i valori della variabile a . Si dica se $H = (1/4, 3/4)$ è un punto di equilibrio. Si spieghi perché lo è o non lo è, e in quest'ultimo caso si individuino gli eventuali punti di equilibrio.

Soluzione

Il punto H non è di equilibrio, perché non è punto di massimo per il giocatore di riga e di minimo per quello di colonna. Lo è invece il punto $K = (4/17, 11/17)$.

⁸Questo esercizio riguarda un gioco continuo. Nel corso si è solo accennato ai giochi nel continuo: per i giochi a somma nulla, la proprietà degli equilibri di essere punti di massimo per il giocatore di riga e minimo per quello di colonna si estende banalmente.

Esercizio 10⁹

In una nazione esistono tre partiti, più o meno della stessa forza: la sinistra (S), il centro (C) e la destra (D). Nel formulare il proprio programma ciascuna formazione può scegliere tra due alternative: un programma conservatore (Cons.) e un programma riformista (Rif.). I programmi vengono resi pubblici contemporaneamente e il blocco di due o tre partiti che propone lo stesso tipo di programma vince le elezioni e forma la maggioranza di governo. Ciascuna delle tre formazioni riceve un beneficio in termini di seggi elettorali se fa parte della maggioranza, mentre paga un costo se si trova isolata all'opposizione. Il beneficio per le formazioni con il programma vincente è +30 se il programma è riformista, +10 se è conservatore. Il danno per la formazione (eventuale) con il programma perdente è -50 se il programma è riformista, -10 se è conservatore.

Si descriva questo problema come gioco in forma strategica. Siccome i giocatori sono tre anziché due, non sarà possibile riportare i valori nella classica tabella, ma bisognerà elencare sulle righe le strategie dei tre giocatori e sulle colonne i corrispondenti *payoff*.

Si dica se esistono punti di equilibrio e quali sono.

Si indichi la strategia consigliata dal criterio del caso pessimo.

Soluzione

La matrice dei *payoff* è la seguente.

x_S	x_C	x_D	$f_S(x_S, x_C, x_D)$	$f_C(x_S, x_C, x_D)$	$f_D(x_S, x_C, x_D)$
Cons.	Cons.	Cons.	10	10	10
Cons.	Cons.	Rif.	10	10	-50
Cons.	Rif.	Cons.	10	-50	10
Cons.	Rif.	Rif.	-10	30	30
Rif.	Cons.	Cons.	-50	10	10
Rif.	Cons.	Rif.	30	-10	30
Rif.	Rif.	Cons.	30	30	-10
Rif.	Rif.	Rif.	30	30	30

I punti di equilibrio sono le terne omogenee di strategie (Cons.,Cons.,Cons.) e (Rif.,Rif.,Rif.). Infatti, in tale situazione ciascun giocatore evita di cambiare strategia perché finirebbe nella minoranza e sarebbe penalizzato. Invece, nelle situazioni miste il giocatore in minoranza è incentivato a unirsi alla maggioranza. Nel caso misto con programma di maggioranza conservatore, inoltre, i giocatori di maggioranza sono incentivati a cambiare alleanza.

La strategia del caso pessimo suggerisce il programma conservatore, dato che il caso pessimo consiste nel finire in minoranza e in tal caso il programma conservatore fa perdere di meno.

Esercizio 11¹⁰

In attesa dell'arrivo dello sceriffo, i due aiutanti *A* e *B* giocano un poker semplificato con un mazzo di tre sole carte: J, Q, K. All'inizio del gioco ciascun giocatore mette

⁹Questo esercizio considera un gioco a tre persone. A parte questo, i concetti di equilibrio e di strategia nel caso pessimo sono identici.

¹⁰Questo esercizio riguarda concetti avanzati, non richiesti per l'esame, dato che include una condizione di incertezza. L'estensione si basa su concetti esposti nei relativi capitoli.

sul tavolo 1 dollaro di piatto. Poi i giocatori prendono una carta a testa, lasciando da parte la terza (ognuno conosce solo la propria carta). A sceglie se “aprire”, cioè aggiungere un altro dollaro alla posta sul piatto, o “passare”. B sceglie se passare, aprire (solo se A non l’ha già fatto) o “vedere” (solo se A ha già aperto). Infine, A può passare o vedere. Riassumendo, il gioco può svolgersi secondo lo schema seguente:

1. A passa:

- (a) B passa: il piatto viene ridiviso fra i due giocatori;
- (b) B apre, cioè aggiunge un terzo dollaro:
 - i. A passa: B vince il piatto, cioè riprende i suoi soldi più 1 dollaro di A ;
 - ii. A vede: il giocatore con la carta più alta vince il piatto, cioè riprende il suo dollaro più 1 dollaro dell’altro giocatore;

2. A apre, cioè aggiunge un terzo dollaro:

- (a) B passa: A vince il piatto, cioè riprende i suoi soldi più 1 dollaro di B ;
- (b) B vede: il giocatore con la carta più alta vince il piatto, cioè riprende il suo dollato più 1 dollaro dell’altro giocatore.

Si elenchino le possibili strategie dei due giocatori.

Si descriva la matrice dei risultati del primo giocatore per ogni possibile distribuzione di carte.

Si descriva la matrice dei valori attesi dei *payoff*.

Si elenchino le possibili strategie per il primo giocatore, cioè le sue possibili giocate in funzione della carta ricevuta inizialmente (ad esempio, la terna $[P - P, P - V, A]$ indicherebbe la strategia di passare sempre se la carta in mano è J o Q, di aprire se è K).

Soluzione

Le strategie possibili per il primo giocatore sono $P - P$, $P - V$ ed A , quelle per il secondo sono P , A e V .

A seconda delle carte estratte dai due giocatori, la matrice dei *payoff* per il primo è la seguente (il simbolo ‘-’ marca i profili di strategie impossibili):

- se la sua carta è J e per qualsiasi carta del secondo giocatore, oppure se la sua carta è Q e quella del secondo giocatore è K:

f	P	A	V
$P - P$	0	-1	-
$P - V$	0	-2	-
A	1	-	-2

- se la sua carta è Q e quella del secondo giocatore è J:

f	P	A	V
$P - P$	0	-1	-
$P - V$	0	2	-
A	1	-	2

- se la sua carta è K e per qualsiasi carta del secondo giocatore:

f	P	A	V
$P - P$	0	-1	-
$P - V$	0	2	-
A	1	-	0

La matrice dei *payoff* attesi è:

f	P	A	V
$P - P$	0	-1	-
$P - V$	0	0	-
A	1	-	0

Il primo giocatore ha $3^3 = 27$ possibili strategie, dato che vi sono 3 possibili mosse per ognuna delle 3 possibili carte iniziali.

Esercizio 12¹¹

Il “gioco della verità” si serve di un’urna contenente due palline (una rossa, R, e una nera, N). Il giocatore A estrae dall’urna una pallina senza mostrarla all’avversario, e fa una dichiarazione, vera o falsa, sul colore di questa. Se dice che la pallina è nera, B vince 1 punto e A ne perde 1. Se dice che è rossa, il giocatore B può accettare la dichiarazione o rifiutarla. Se la accetta, A vince 1 punto e B ne perde 1. Infine, se la rifiuta, si controlla il colore della pallina: se è rossa, A vince 2 punti e B ne perde 2; se è nera, B vince 2 punti e A ne perde 2. Si suppone che le probabilità di estrazione delle due palline siano equilibrate (50% per una).

Si descriva il gioco in forma strategica, attraverso una tabella dei risultati.

Si calcolino graficamente le strategie ottime per entrambi i giocatori.

Si dica come cambia il gioco e come cambiano le strategie ottime se l’urna contiene una pallina rossa e due nere.

Soluzione

La difficoltà principale dell’esercizio sta nel distinguere ciò che ricade nello scenario da ciò che ricade nella strategia. Il colore della pallina estratta è lo scenario. Per il giocatore A la strategia può essere:

1. dichiarare il colore esatto della pallina;
2. dichiarare che la pallina è rossa, qualunque sia il colore.

Sarebbe più spontaneo distinguere fra dichiarare il vero e il falso, ma la strategia di dire il falso è vietata nel caso di pallina rossa, mentre una descrizione corretta del gioco richiede che le strategie siano tutte lecite in ogni scenario.

La strategia per il giocatore B può essere:

1. accettare la dichiarazione;

¹¹ Anche questo esercizio riguarda un gioco in condizioni di incertezza. Chi trovasse l’esercizio o la soluzione poco chiari, sarà forse consolato da un aneddoto: questo esercizio, scritto in una forma nettamente più ambigua, mi venne sottoposto da studente in un compitino. Lo fraintese completamente, perdendo 6 punti. Fu così forte il senso di essere stato defraudato che decisi di rifiutare il voto; all’appello successivo presi 30 e lode.

2. rifiutare la dichiarazione.

Nel caso la pallina sia rossa, la tabella dei *payoff* è:

f	Accettare	Rifiutare
Vero	1	2
Rossa	1	2

Nel caso la pallina sia nera, la tabella dei *payoff* è:

f	Accettare	Rifiutare
Vero	-1	-1
Rossa	1	-2

La tabella dei *payoff* attesi è quindi:

f	Accettare	Rifiutare
Vero	0	$1/2$
Rossa	1	0

La strategia ottima per il giocatore A è $(2/3, 1/3)$, quella per il giocatore B è $(1/3, 2/3)$. Il risultato atteso è una vittoria di $1/3$ per A .

Il caso modificato produce la tabella

f	Accettare	Rifiutare
Vero	$-1/3$	0
Rossa	0	$-2/3$

con strategie ottime $(5/6, 1/6)$ per A e $(1/3, 2/3)$ per B . Il risultato atteso è una vittoria di $-1/9$ per A .

Capitolo 15

Teoria delle decisioni di gruppo

Anche nella teoria delle decisioni di gruppo è l'insieme di decisori D è finito, ma non ridotto a un solo elemento ($|D| > 1$). La differenza fondamentale rispetto alla teoria dei giochi è che i vari decisori non fissano indipendentemente il valore delle variabili di decisione, ciascuno per il sottoinsieme che gli compete, ma devono scegliere l'alternativa da realizzare concordemente, coordinandosi in qualche modo. Una situazione tipica è quella in cui esiste un decisore principale che indica la soluzione, ma lo deve fare in modo da soddisfare il più possibile le preferenze degli altri, oltre alla propria.

Un approccio classico alla modellazione di questi problemi è la costruzione, a partire dalle preferenze Π_d dei singoli decisori $d \in D$, di una singola relazione di preferenza $\Pi_D \subseteq F \times F$, detta *preferenza di gruppo*, che esprime i criteri in base ai quali la decisione comune dovrà essere presa. Una volta costruita tale relazione, il problema diventa equivalente a quello di un decisore singolo, e gli si applicano tutte le tecniche descritte nei capitoli precedenti. Il passaggio fondamentale della risoluzione, quindi, è il modo in cui si costruisce una relazione di preferenza a partire da $|D|$ relazioni date. Gli studi su questo problema si riferiscono soprattutto alle scienze politiche e sociali, e ci si può quindi aspettare che la nomenclatura, ancora una volta, cambi. In effetti, i decisori vengono spesso definiti *giudici*, *individui*, *elettori*, *agenti* ovvero *cittadini*.

Nel seguito, assumeremo che:

- vi sia un solo scenario possibile ($|\Omega| = 1$), condizione introdotta per semplicità e anche perché spesso nelle decisioni di gruppo ognuno ragiona in base a una propria opinione abbastanza precisa (giusta o sbagliata che sia) sull'impatto che ciascuna soluzione avrà;
- l'insieme delle soluzioni X sia finito;
- la funzione impatto $f(x)$ sia invertibile, cioè ogni soluzione abbia il proprio impatto, e quindi le relazioni di preferenza si possano pensare come relazioni su X , anziché su F (basta definire $x \preceq x'$ se e solo se $f(x) \preceq f(x')$);
- le relazioni di preferenza dei singoli decisori siano delle relazioni di ordine debole, cioè riflessive, transitive e complete.

L'ultima condizione è la più importante. Come sappiamo, significa che si richiedono delle condizioni di razionalità un po' più stretta a ciascuno dei decisori.

Definizione 57 Dato un insieme X di soluzioni, definiamo $\mathcal{D}(X)$ l'insieme di tutti gli ordini deboli sugli elementi di X .

Siccome ogni relazione di preferenza fra soluzioni è un insieme di coppie di soluzioni ($\Pi_d \subseteq X \times X$ ovvero $\Pi_d \in 2^{X \times X}$), l'insieme $\mathcal{D}(X)$ raccoglie le relazioni che godono delle proprietà riflessiva, transitiva e completa, e dunque $\mathcal{D}(X) \subset 2^{X \times X}$.

Esempio 86 Consideriamo un insieme di tre soluzioni, $X = \{a, b, c\}$. L'insieme degli ordini deboli su X comprende 19 diverse relazioni:

- i 6 ordini totali, vale a dire le permutazioni delle tre soluzioni:

$$\begin{array}{ll} a \prec b \prec c & a \prec c \prec b \\ b \prec a \prec c & b \prec c \prec a \\ c \prec a \prec b & c \prec b \prec a \end{array}$$

- i 6 ordini deboli in cui due soluzioni sono indifferenti fra loro e preferite alla terza:

$$\begin{array}{ll} a \sim b \prec c & a \sim c \prec b \\ b \sim a \prec c & b \sim c \prec a \\ c \sim a \prec b & c \sim b \prec a \end{array}$$

- i 6 ordini deboli in cui una soluzione è preferita alle altre due, che sono indifferenti fra loro:

$$\begin{array}{ll} a \prec b \sim c & a \prec c \sim b \\ b \prec a \sim c & b \prec c \sim a \\ c \prec a \sim b & c \prec b \sim a \end{array}$$

- l'ordine debole in cui le tre soluzioni sono tutte indifferenti fra loro:

$$a \sim b \sim c$$

Ciascuno di questi ordini deboli si può esprimere (con una certa pesantezza) elencando le coppie di soluzioni tali che la prima è debolmente preferita alla seconda. Per esempio, l'ordine $a \prec b \prec c$ si può esprimere come $\{(a, a), (a, b), (a, c), (b, b), (b, c), (c, c)\}$, mentre l'ordine $a \sim b \sim c$ si può esprimere come $\{(a, a), (a, b), (a, c), (b, a), (b, b), (b, c), (c, a), (c, b), (c, c)\}$.

15.1 Costituzione

Il problema che ci poniamo è come costruire una relazione di preferenza di gruppo a partire da un insieme finito di ordini deboli, uno per ogni decisore.

Definizione 58 Si dice costituzione o funzione di benessere collettivo una funzione che associa ad ogni $|D|$ -upla di ordini deboli su un insieme di soluzioni dato un ordine debole (“di gruppo”) sullo stesso insieme:

$$g : \mathcal{D}(X)^{|D|} \rightarrow \mathcal{D}(X)$$

Se si procede in modo arbitrario, si possono ovviamente costruire moltissime funzioni che riducano molti ordini deboli a uno solo. L'aspetto interessante è cercare una funzione che rispetti il più possibile gli ordini deboli di partenza. Per fare questo, tale funzione deve godere di opportune proprietà. Nel seguito, passiamo in rassegna alcune proposte avanzate via via nel tempo, e ne studiamo i limiti. Quindi, enunceremo le proprietà che sembrano necessarie ad una costituzione accettabile, e dimostreremo il fondamentale risultato secondo il quale non esistono costituzioni che rispettino tutte le proprietà necessarie. Infine, discuteremo le critiche a tale risultato e alcuni sviluppi successivi degli studi sull'argomento.

15.2 Metodo di Condorcet

Il *metodo di Condorcet*¹, anche detto *metodo delle maggioranze semplici* si basa su questa definizione:

$$x \preceq_D x' \Leftrightarrow |\{d \in D : x \preceq_d x'\}| \geq |\{d \in D : x' \preceq_d x\}|$$

cioè una soluzione è preferibile a un'altra secondo il gruppo quando il numero di decisori che preferisce la prima alla seconda supera il numero dei decisori che preferisce la seconda alla prima.

Si noti che i decisori secondo i quali le due soluzioni sono indifferenti compaiono a entrambi i membri, e quindi si elidono: contano solo quelli che hanno una preferenza stretta. Di conseguenza, due soluzioni sono indifferenti quando lo stesso numero di decisori preferisce la prima alla seconda e la seconda alla prima (in particolare, quando tutti i decisori le considerano indifferenti).

Il metodo di Condorcet fornisce un semplice algoritmo per calcolare una funzione di benessere collettivo. Tuttavia, ha un grosso difetto, che è noto in letteratura come *paradosso di Condorcet*. Il difetto è illustrato dalla Tabella 15.1: vi sono tre decisori $D = \{1, 2, 3\}$ e tre soluzioni $X = \{a, b, c\}$. Ogni colonna è associata a un decisore e descrive la sua relazione di preferenza sulle soluzioni: nella riga 1 è riportata la soluzione migliore, nella riga 2 la seconda e nella riga 3 la peggiore. Le tre relazioni di preferenza individuale sono quindi degli ordini totali, senza *ex aequo*.

Ordine	Decisori		
	1	2	3
1	a	b	c
2	b	c	a
3	c	a	b

Tabella 15.1: Esempio del paradosso di Condorcet: in base al metodo di Condorcet, ciascuna delle tre soluzioni è migliore di un'altra; la relazione di preferenza risultante non è transitiva, dunque non è un ordine debole.

Applicando il metodo di Condorcet alla tabella, si ottiene che:

- $a \prec_D b$ perché a è preferibile a b per due decisori (1 e 3), mentre il contrario vale solo per uno (2);
- $b \prec_D c$ perché b è preferibile a c per due decisori (1 e 2), mentre il contrario vale solo per uno (3);

¹Marie Jean Antoine Nicolas de Caritat, marchese di Condorcet (1743-1794), matematico, economista, filosofo e politico attivo durante la rivoluzione francese nel partito dei girondini, suicida in carcere durante il periodo del Terrore.

- $c \prec_D a$ perché c è preferibile ad a per due decisori (2 e 3), mentre il contrario vale solo per uno (1).

Inoltre, banalmente ciascuna soluzione è preferibile a sé stessa perché lo è per tutti i decisori. La relazione risultante presenta una preferenza stretta circolare:

$$a \prec b \prec c \prec a$$

e non è transitiva, dato che $a \prec b$ e $b \prec$, ma $a \not\prec c$. Quindi, *il metodo di Condorcet non sempre consente di costruire un ordine debole*. Questo non significa che non funziona mai, ma che può non funzionare. E che, in particolare, può non fornire una soluzione preferibile a tutte le altre, che è il requisito minimo per un criterio di scelta.

La storia presenta diversi esempi di voti parlamentari nei quali tre soluzioni legislative avevano a sostegno gruppi di decisori strutturati come nel paradosso di Condorcet. In alcuni casi, l'*impasse* legislativa venne risolta introducendo un ordine dei lavori arbitrario, cioè discutendo prima il confronto fra le soluzioni a e b (da cui uscì vincitrice la soluzione b) e poi il confronto fra la soluzione vincente b e la terza soluzione c (da cui uscì vincitrice la soluzione c). Un altro ordine dei lavori avrebbe condotto a una scelta diversa. Chi si occupa di politica sa che fissare l'ordine dei lavori è spesso in una posizione di forza nei lavori parlamentari.

15.3 Metodo di Borda

Il *metodo di Borda*² si basa su una definizione ausiliaria che già abbiamo introdotto nella Sezione ??, quella del *conto di Borda*:

$$B_d(x) = |\{x' \in X : x \preceq_d x'\}|$$

cioè il valore di una soluzione x per il decisore d è dato dal numero di soluzioni alle quali il decisore preferisce x . Una volta introdotta una funzione valore per i singoli decisori, aggregare queste funzioni con una semplice somma appare uno sviluppo naturale, in una prospettiva ugualitaria e democratica:

$$B_D(x) = \sum_{d \in D} B_d(x)$$

e altrettanto naturale è basare sulla funzione aggregata, cioè sul conto di Borda complessivo, la preferenza di gruppo:

$$x \preceq_D x' \Leftrightarrow B_D(x) \geq B_D(x')$$

cioè una soluzione è preferibile a un'altra secondo il gruppo quando il conto di Borda complessivo su tutti i decisori della prima supera quello della seconda.

Questa definizione introduce una funzione valore conforme per il gruppo e basa su di essa la preferenza. La preferenza che ne risulta non può che essere un ordine debole. Dunque, a differenza del metodo di Condorcet, il metodo di Borda consente sempre di costruire una costituzione.

Anche il metodo di Borda, però, ha un difetto, che abbiamo già incontrato in un altro contesto, cioè la dipendenza da alternative irrilevanti. In altre parole, l'ordine debole che si viene a creare fra due soluzioni non dipende solo dalle preferenze dei

²Jean-Charles de Borda (1733-1799), matematico, fisico e ammiraglio francese attivo durante la rivoluzione francese.

decisori fra le due soluzioni, ma in generale anche dalla preferenze che essi hanno rispetto ad altre soluzioni. Questo è grave perché in generale l'insieme delle soluzioni costruito nella fase di modellazione è almeno in parte arbitrario (se n'è discusso nella Sezione 2.1.2, accennando al procedimento iterativo di costruzione del modello, e in particolare dell'insieme delle soluzioni fra cui sceglierne; per esempio, le opere compensative per grandi interventi pubblici sono tipicamente definite in fasi successive dello studio, dopo aver identificato e discusso gli impatti dell'intervento in sé). Se il risultato finale della decisione dipende dalle alternative prese in considerazione, si aprono spazi di manipolazione del procedimento, che non sono desiderabili.

Esempio 87 Si consideri il problema di decisione rappresentato dalla Tabella 15.2, caratterizzato da sette decisori ($D = \{1, 2, 3, 4, 5, 6, 7\}$) e quattro soluzioni ($X = \{a, b, c, d\}$).

Ordine	Decisori						
	1	2	3	4	5	6	7
1	a	b	c	a	b	c	a
2	b	c	d	b	c	d	b
3	c	d	a	c	d	a	c
4	d	a	b	d	a	b	d

Tabella 15.2: Un problema di decisione di gruppo con 7 decisori e 4 soluzioni; ogni colonna è associata a un decisore e riporta dalla prima alla quarta riga le soluzioni in ordine (totale) di preferenza.

Anche in questo caso, le relazioni di preferenza di tutti i decisori sono ordini totali: alla soluzione preferita viene associato un conto di Borda pari a $B_d(x) = 4$, alla seconda un valore pari a 3, alla terza pari a 2 e all'ultima pari a 1. Il conto complessivo è riportato nella Tabella 15.3, da cui si ricava l'ordinamento totale $c \prec b \prec a \prec d$.

X	$B_D(x)$
a	$4 + 1 + 2 + 4 + 1 + 2 + 4 = 18$
b	$3 + 4 + 1 + 3 + 4 + 1 + 3 = 19$
c	$2 + 3 + 4 + 2 + 3 + 4 + 2 = 20$
d	$1 + 2 + 3 + 1 + 2 + 3 + 1 = 13$

Tabella 15.3: Conto complessivo di Borda per le alternative del problema riportato in Tabella 15.2.

Se però si cancellasse la soluzione d , che appare piuttosto inutile, dato che nessun decisore la predilige e in complesso ha un risultato nettamente inferiore alle altre, si otterrebbero il problema e il conto complessivo riportati, rispettivamente, nelle Tabelle 15.4 e 15.5.

Ordine	Decisori						
	1	2	3	4	5	6	7
1	a	b	c	a	b	c	a
2	b	c	a	b	c	a	b
3	c	a	b	c	a	b	c

Tabella 15.4: Il problema della Tabella 15.2 ridotto escludendo la soluzione d .

X	$B_D(x)$
a	$3 + 1 + 2 + 3 + 1 + 2 + 3 = 15$
b	$2 + 3 + 1 + 2 + 3 + 1 + 2 = 14$
c	$1 + 2 + 3 + 1 + 2 + 3 + 1 = 13$

Tabella 15.5: Conto complessivo di Borda per le alternative del problema riportato in Tabella 15.4.

Dal conto complessivo si ricava l'ordinamento totale $a \prec b \prec c$, che ovviamente non include l'alternativa eliminata, ma inverte completamente la relazione d'ordine fra le altre tre alternative. Il motivo di questo cambiamento è che il conto di Borda di ciascuna alternativa, e quindi la relazione di preferenza fra due alternative, dipende dal posizionamento di tutte le altre. In particolare, in questo caso l'alternativa d era sempre peggiore di c , spesso peggiore di b e solo talvolta di a ; eliminarla ha danneggiato molto il conto di d , abbastanza quello di b e poco quello di a , invertendo le relazioni di preferenza fra queste tre alternative.

L'esempio precedente porta a concludere che anche il metodo di Borda può non essere soddisfacente, dato che è aperto a manipolazioni nella fase di scelta delle alternative da confrontare.

15.4 Metodo lessicografico

Il metodo lessicografico perde una caratteristica fondamentale dei due precedenti, cioè la simmetria rispetto ai singoli decisori. In effetti, il metodo consiste nello scegliere una permutazione dei decisori stessi, $p = (d_1, d_2, \dots, d_{|D|})$ e definire la preferenza fra due soluzioni come la preferenza del primo decisore secondo la permutazione che non sia indifferente fra loro. Due alternative sono indifferenti, quindi, se e solo se tutti i decisori le considerano tali.

$$x \prec_D x' \Leftrightarrow \begin{cases} x \prec_{d_1} x' \text{ oppure} \\ x \sim_{d_1} x' \text{ e } x \prec_{d_2} x' \text{ oppure} \\ \dots \\ x \sim_{d_1} x', x \sim_{d_2} x', \dots \text{ e } x \prec_{d_{|D|}} x' \end{cases}$$

$$x \sim_D x' \Leftrightarrow x \sim_{d_1} x', x \sim_{d_2} x', \dots \text{ e } x \sim_{d_{|D|}} x'$$

Questo metodo descrive una versione estrema di una monarchia assoluta, nella quale la volontà del monarca d_1 è legge (per lo meno nella scelta sull'insieme X : su altre scelte, l'ordine potrebbe cambiare) e, se il monarca è indifferente, delega la scelta a un altro, che eventualmente la delega a un terzo se è indifferente, e così via sino all'ultimo cittadino del regno, che può esprimere una preferenza se e solo se tutti gli altri sono indifferenti.

15.5 Sistema di pluralità

Il sistema di pluralità è semplicemente il metodo usato nelle elezioni più tradizionali: ogni decisore riassume la propria preferenza nell'indicare una soluzione. La soluzione che ha più voti vince. Rigorosamente parlando, dato che ipotizziamo che le relazioni di preferenza individuali siano ordini deboli, riassumiamo la preferenza di un decisore nell'indicare tutte le alternative (fra loro indifferenti) che sono

strettamente preferibili alle altre. In termini formali, si definisce:

$$V(x) = |\{d \in D : x \preceq_d y \text{ per ogni } y \in X\}|$$

e

$$x \preceq_D x' \Leftrightarrow V(x) \geq V(x')$$

Come il metodo di Borda, questo sistema si basa su una funzione valore conforme, e quindi dà luogo a un ordine debole. Tuttavia ha anch'esso il difetto di dipendere dalle alternative irrilevanti, anche se un po' meno del metodo di Borda, dato che la funzione valore non conteggia tutte le posizioni di ogni soluzione, ma solo le posizioni vincenti. Per cambiare il numero di posizioni vincenti, occorre introdurre o eliminare alternative scelte in modo abbastanza delicato.

Esempio 88 Si consideri il problema di decisione di gruppo con 4 soluzioni e 7 decisori rappresentato nella Tabella 15.6.

Ordine	Decisori						
	1	2	3	4	5	6	7
1	a	a	b	b	d	d	d
2	b	b	a	a	a	b	c
3	c	c	c	c	b	c	a
4	d	d	d	d	c	a	b

Tabella 15.6: Un problema di decisione di gruppo con 7 decisori e 4 soluzioni.

Ancora una volta, le relazioni di preferenza dei decisori sono tutte ordini totali. Il metodo del sistema di pluralità porta a scegliere la soluzione *d*, perché è la preferita in assoluto per tre decisori, mentre *a* e *b* sono le preferite per due e *d* non lo è per nessuno. Quindi, *d* \prec *a* \sim *b* \prec *c*.

Tuttavia, se si eliminasse la soluzione *b*, riducendosi al problema rappresentato nella Tabella 15.7, il metodo indicherebbe la soluzione *a*, in quanto preferita da quattro decisori, contro i tre della soluzione *d* e nessuno per la soluzione *c*. Quindi *a* \prec *d* \prec *c*.

Ordine	Decisori						
	1	2	3	4	5	6	7
1	a	a	a	a	d	d	d
2	c	c	c	c	a	c	c
3	d	d	d	d	c	a	a

Tabella 15.7: Il problema della Tabella 15.2 ridotto escludendo la soluzione *b*.

In realtà, il metodo ha un altro difetto abbastanza grave: una soluzione fortemente gradita a una minoranza può prevalere su una soluzione che non è preferita da nessuno, ma che piace abbastanza a tutti. Questo deriva dal fatto di considerare nella funzione $V(x)$ solo la prima posizione negli ordini deboli dei decisori, ignorando del tutto le successive.

Esempio 89 Considerando il problema con quattro soluzioni e sette decisori rappresentato nella Tabella refTab:EsempioPluralita3, si può osservare che la soluzione vincente *d* è molto sgradita agli altri decisori, mentre la soluzione *a*, che nessuno sostiene in prima posizione, è però in seconda posizione per tutti, e sarebbe quindi probabilmente un buon compromesso. Un metodo che ignori le soluzioni in posizioni successive alla prima non può rendersi conto dell'esistenza di queste situazioni.

Ordine	Decisori						
	1	2	3	4	5	6	7
1	b	b	c	c	d	d	d
2	a	a	a	a	a	a	a
3	c	c	b	b	b	c	c
4	d	d	d	d	c	b	b

Tabella 15.8: Un altro problema di decisione di gruppo con 7 decisori e 4 soluzioni.

15.6 L'approccio assiomatico

A questo punto, dopo aver passato in analisi una serie di metodi abbastanza ragionevoli e naturali, ma tutti affetti da punti deboli, ci si può chiedere se esista un metodo per aggregare le preferenze dei singoli in una preferenza di gruppo rispettando alcune proprietà desiderabili di fondo. E si può ipotizzare che elencare tali proprietà sia il primo passo per costruire tale metodo. Questo è l'approccio assiomatico proposto da Arrow³ nel 1951, con un esito inatteso, cioè la dimostrazione che è impossibile formulare un metodo che goda di tutte le proprietà considerate fondamentali per aggregare le preferenze in modo democraticamente accettabile. Per descrivere in modo compatto il risultato di Arrow, è utile dare un nome a ciascun possibile insieme di preferenze che una costituzione dovrebbe aggregare.

Definizione 59 *Dato un insieme X finito di alternative e un insieme D finito di decisori, si dice profilo $\mathcal{P}(X)$ qualsiasi $|D|$ -upla di ordini deboli fra le alternative associati ai decisori.*

Il problema che stiamo affrontando consiste quindi nel determinare una costituzione g che aggreghi qualsiasi possibile profilo $\mathcal{P}(X) \in \mathcal{D}(X)^{|D|}$ in una relazione di preferenza di gruppo $\Pi_D(X)$ che sia un ordine debole:

$$\Pi_D(X) = g(\mathcal{P}(X)) \in D(X) \text{ per ogni } \mathcal{P} \in \mathcal{D}(X)^{|D|}$$

Dati due profili e due alternative, è interessante osservare se uno dei due risulta più sbilanciato dell'altro a favore di una delle due alternative. Per visualizzare la situazione si può pensare a una prima indagine, durante la quale si registra il profilo delle preferenze dei decisori fra le due alternative, dopo la quale avviene un evento che convince alcuni dei sostenitori della seconda alternativa a passare alla prima. A questo punto, una seconda indagine registra il nuovo profilo, che ovviamente è più sbilanciato del primo a favore della prima alternativa.

Definizione 60 *Date due alternative $x, y \in X$ e due profili $\mathcal{P}(X)$ e $\mathcal{P}'(X)$, si indicano con \prec_d e \prec'_d le preferenze strette del decisore $d \in D$ nei profili $\mathcal{P}(X)$ e $\mathcal{P}'(X)$. Analogamente si distinguono l'indifferenza (\sim_d e \sim'_d) e la preferenza debole (\preceq_d e \preceq'_d) nei due profili. Inoltre, si dice che $\mathcal{P}'(X)$ promuove x contro y più di $\mathcal{P}(X)$ quando*

- ogni decisore strettamente favorevole a x in $\mathcal{P}(X)$ rimane strettamente favorevole a x in $\mathcal{P}'(X)$:

$$x \prec_d y \Rightarrow x \prec'_d y$$

³Kenneth Joseph Arrow (1921 - 2017), economista statunitense, vincitore del Premio Nobel per l'Economia nel 1972.

- ogni decisore indifferente fra x e y in $\mathcal{P}(X)$ è almeno debolmente favorevole a x in $\mathcal{P}'(X)$:

$$x \sim_d y \Rightarrow x \preceq'_d y$$

Infine, occorre definire il concetto di dittatura, per poterlo escludere. Un dittatore è un decisore tale che qualunque sua preferenza stretta si traduce in una preferenza stretta dell'intero gruppo. Nel metodo lessicografico, il decisore d_1 è un dittatore. Gli altri metodi considerati, invece, non ammettono dittatori, nel senso che si può sempre pensare a un profilo nel quale almeno una delle preferenze strette di ciascun decisore viene violata dal gruppo.

Definizione 61 Data una costituzione g , un dittatore in tale costituzione è un decisore $d \in D$ tale che per qualsiasi profilo in cui $x \prec_d y$, la costituzione impone che $x \prec_D y$.

Esiste infine una nozione limitata di dittatura, che viene sfruttata nel teorema di Arrow: anziché avere un solo decisore le cui preferenze strette determinano quelle del gruppo, si può avere un sottoinsieme di decisori la cui preferenza stretta su una specifica coppia di soluzioni, se sono unanimi, determina quella del gruppo.

Definizione 62 Data una costituzione g , si dice che in essa un sottoinsieme di decisori $V \subseteq D$ è decisivo per una coppia di soluzioni (x, y) quando, per ogni profilo $\mathcal{P}(X)$ tale che $x \prec_d y$ per ogni $d \in V$, la costituzione impone $x \prec_D y$.

Un dittatore è ovviamente un caso particolare di insieme decisivo, costituito da un solo decisore e valido per qualsiasi coppia di soluzioni.

L'approccio di Arrow parte introducendo alcuni assiomi che definiscono le proprietà richieste a una costituzione $g : \mathcal{D}(X)^{|D|} \rightarrow \mathcal{D}(X)$ per potersi dire democratica. Quindi, dimostra che, se tutti gli assiomi tranne l'ultimo sono veri, necessariamente l'ultimo è falso. La dimostrazione originaria di Arrow considera sette assiomi.

Teorema 26 (Teorema di Arrow) Siano dati un insieme X finito di alternative, un insieme D finito di decisori e una relazione di preferenza $\Pi(X) : D \rightarrow \mathcal{D}(X)$ che associa ad ogni decisore un ordine debole sull'insieme delle alternative. I seguenti sette enunciati sono tra loro incompatibili:

1. *assioma di non banalità:* ci sono almeno tre alternative e almeno due decisori
 $|X| \geq 3 \quad |D| \geq 2$
2. *assioma del dominio universale:* esiste una costituzione g definita su tutti i possibili profili, cioè g descrive come aggregare qualsiasi insieme di ordini deboli associati ai decisori;
3. *assioma dell'ordine debole:* la costituzione g dà luogo a un ordine debole (relazione di preferenza riflessiva, transitiva e completa) su X ;
4. *assioma di indipendenza dalle alternative irrilevanti:* dato un problema con un insieme ridotto di alternative $X' \subset X$, gli stessi decisori e relazioni di preferenza limitate alle alternative residue, la costituzione dà luogo allo stesso ordine debole che si otterrebbe limitando $\Pi_D(X)$ all'insieme ridotto X' .
5. *assioma di monotonia:* se un profilo $\mathcal{P}'(X)$ promuove x contro y più di $\mathcal{P}(X)$, se per tutte le coppie di soluzioni che non includono x i due profili sono identici e se $x \prec_D y$ (cioè il gruppo preferisce x nel profilo $\mathcal{P}(X)$), allora $x \prec'_D y$ (cioè il gruppo preferisce x anche in $\mathcal{P}'(X)$);

6. assioma di sovranità popolare: per ogni coppia di soluzioni (x, y) , deve esistere almeno un profilo tale che $x \preceq_D y$.
7. assioma di non dittatura: non esiste alcun dittatore nella costituzione g .

Dimostrazione. La dimostrazione attraversa le seguenti fasi:

1. dimostra l'esistenza di un insieme decisivo per una coppia di soluzioni;
2. dimostra che tale insieme è riducibile a un singolo decisore;
3. dimostra che tale decisore è decisivo per ogni coppia di soluzioni, e quindi è un dittatore.

Per prima cosa, osserviamo che l'assioma di non banalità garantisce l'esistenza di almeno tre soluzioni. Quindi, osserviamo che l'assioma di monotonia e quello di sovranità popolare implicano la seguente proprietà di *unanimità* (o assioma di Pareto): se $x \prec_d y$ per ogni $d \in D$, allora $x \prec_D y$, cioè la costituzione impone che le preferenze strette unanimi si traducano in preferenze strette del gruppo. Infatti, se esiste un profilo che fa vincere x su y e tutti gli altri profili che promuovono x più di quello fanno anch'essi vincere x , allora il profilo in cui tutti sono unanimi nel preferire x farà ancora vincere x . Non è invece vero il contrario.

Grazie all'assioma di unanimità, in qualsiasi costituzione ogni coppia di soluzioni (x, y) ammette un insieme decisivo. Infatti, se tutti i decisori di D , unanimi, preferiscono strettamente x a y , il gruppo preferisce strettamente x a y . Questo implica che D rispetti la definizione di insieme decisivo per la coppia (x, y) , qualunque essa sia.

Ora viene il punto più delicato dell'intera dimostrazione: la ricerca di un insieme decisivo minimale, cioè un insieme che non può perdere nessun membro senza perdere la qualifica di insieme decisivo. Assumiamo che $V \subseteq D$ sia un insieme decisivo minimale. Questo significa che è decisivo per una o più coppie di soluzioni, e che privato di qualsiasi suo membro non è più decisivo per nessuna coppia di soluzioni. Quanti elementi ha V ?

Chiamiamo d un decisore di V e W l'insieme ottenuto privando V di d , cioè $W = V \setminus \{d\}$. Infine, indichiamo con $D \setminus V$ l'insieme dei decisori che non fanno parte di V ; questo insieme potrebbe anche essere vuoto. Inoltre, chiamiamo (x, y) una qualsiasi delle coppie per cui V è decisivo. La proprietà di essere decisivo è una proprietà che vale per ogni profilo di preferenze: qualunque opinione abbiamo i decisori, se quelli di V sono unanimi nel preferire x a y , allora il gruppo preferisce x a y anche se tutti gli altri decisori la pensano esattamente al contrario. Allora consideriamo la situazione descritta nella tabella seguente, dove ogni colonna riporta l'ordine delle preferenze di un sottoinsieme di decisori rispetto a tre alternative, indicate con x , y e z (qui viene usato l'assioma di non banalità). Si tratta di ordini totali, dunque ordini deboli, per cui il profilo è corretto e la costituzione g è in grado di generare una preferenza di gruppo (per l'assioma del dominio universale). Quale sarà la preferenza del gruppo?

$\{d\}$	$V \setminus \{d\}$	$D \setminus V$
x	z	y
y	x	z
z	y	x

I decisori di V sono unanimi nel preferire strettamente x a y . Siccome V è insieme decisivo per (x, y) , ne deriva che il gruppo ha la stessa preferenza:

$$x \prec_V y \Rightarrow x \prec_D y$$

Ora ci sono due casi possibili: o V contiene più di decisore o ne contiene uno solo. Analizziamo il primo caso: $V \setminus \{d\}$ è un insieme non vuoto. Inoltre, non è decisivo per nessuna coppia di soluzioni (in caso contrario negheremmo l'ipotesi iniziale, che V fosse insieme decisivo minimale). In particolare, non è decisivo per (z, y) . Ora osserviamo le preferenze dei vari decisori per la coppia (z, y) : i membri di $V \setminus \{d\}$ sono unanimi nel preferire strettamente z a y , tutti gli altri decisori sono unanimi nell'avere la preferenza contraria. Quale sarà la preferenza del gruppo fra z e y ? Anzi tutto, una preferenza c'è perché, per l'assioma dell'ordine debole, la relazione di preferenza del gruppo è completa. D'altra parte, la preferenza non può essere strettamente a favore di z , dato che in tal caso concorderebbe con quella dei membri di $V \setminus \{d\}$, che quindi risulterebbe decisivo. Ne deriva che $y \preceq_D z$. Ma, siccome $x \prec_D y$ e la preferenza di gruppo è transitiva (ancora l'assioma dell'ordine debole), ne deriva che

$$x \prec_D y \text{ e } y \preceq_D z \Rightarrow x \prec_D z$$

Tuttavia, l'unico decisore che preferisce strettamente x a z è d , e quindi d è decisivo per la coppia (x, z) . Tutto questo nega l'ipotesi che V fosse insieme decisivo minimale. La conclusione è che, dei due casi, l'unico possibile è il secondo, cioè che V contenga un solo decisore. Di conseguenza, esiste un decisore d che è dittatore specifico per almeno una coppia (x, z) di soluzioni.

Nel secondo passo della dimostrazione, proviamo che d è dittatore per tutte le coppie di soluzioni. Cominciamo con le coppie (x, w) dove $w \neq z$. La tabella seguente illustra un possibile profilo di preferenze (si noti che non stiamo assumendo l'esistenza di quattro soluzioni, dato che potrebbe essere $w = z$; abbiamo cambiato nome solo per non suggerire che l'alternativa sia la stessa usata nel passo precedente).

$\{d\}$	$D \setminus \{d\}$
x	z
z	w
w	x

Come si vede dalla tabella, i decisori sono unanimi nel preferire strettamente y a w , e quindi $y \prec_D w$. D'altra parte, essendo d dittatore specifico per la coppia (x, y) , la sua preferenza stretta per x si traduce in preferenza del gruppo, anche se nessun altro la condivide. A questo punto, la transitività colpisce ancora, imponendo che:

$$x \prec_D y \text{ e } y \prec_D w \Rightarrow x \prec_D w$$

ma l'unico decisore che preferisce strettamente x a w è d , che quindi si rivela dittatore anche per questa seconda coppia. E siccome w è del tutto generica, per tutte le coppie con la soluzione x come primo membro.

Ora si può dimostrare che d è dittatore anche per le coppie (y, w) con $y \neq x$ e $w \neq x$. Per farlo, si consideri la tabella seguente:

$\{d\}$	$D \setminus \{d\}$
y	w
x	y
w	x

In essa, i decisori sono unanimi nel preferire strettamente z a x , e quindi $z \prec_D x$. D'altra parte, essendo d dittatore specifico per le coppie (x, w) , la sua preferenza stretta per x si traduce in preferenza del gruppo, anche se nessun altro la condivide. A questo punto, la transitività colpisce ancora, imponendo che:

$$z \prec_D x \text{ e } x \prec_D w \Rightarrow z \prec_D w$$

ma l'unico decisore che preferisce strettamente y a w è d , che quindi si rivela dittatore anche per la coppia (y, w) , che è del tutto generica, salvo escludere la soluzione x da entrambi i membri.

Infine si può dimostrare che d è dittatore anche per le coppie (y, x) con $y \neq x$. Per farlo, si consideri la tabella seguente:

$\{d\}$	$D \setminus \{d\}$
z	y
y	x
x	z

In essa, i decisori sono unanimi nel preferire strettamente y a x , e quindi $y \prec_D x$. D'altra parte, essendo d dittatore specifico per le coppie (z, y) , la sua preferenza stretta per z si traduce in preferenza del gruppo, anche se nessun altro la condivide. A questo punto, la transitività colpisce ancora, imponendo che:

$$z \prec_D y \text{ e } y \prec_D x \Rightarrow z \prec_D x$$

ma l'unico decisore che preferisce strettamente z a x è d , che quindi si rivela dittatore anche per la coppia (z, x) , dove z è del tutto generica, salvo differire da x . Ovviamente, le coppie (x, x) sono indifferenti per tutti i decisori e per il gruppo nel suo complesso, per riflessività.

La conclusione è che d è un dittatore, violando l'ultimo enunciato del teorema.

■ Più recentemente, fu proposta una dimostrazione basata su sei assiomi, più deboli di quelli originali e più semplici da enunciare. La formulazione moderna sostituisce

- al posto dell'assioma di indipendenza dalle alternative irrilevanti un assioma di *rilevanza binaria*: per qualsiasi coppia di soluzioni (x, y) , la preferenza del gruppo dipende solo dalle preferenze dei singoli decisori fra x e y .

Si dimostra che questo assioma equivale a quello di indipendenza dalle alternative irrilevanti, cioè che aggiungere o togliere alternative non modifica la preferenza fra x e y ; quindi la sostituzione non cambia nulla;

- al posto dei due assiomi di monotonia e sovranità popolare, quello di unanimità, che è più debole.

Nonostante la formulazione moderna del teorema consideri un insieme di potenziali costituzioni più ampio di quella originale, anche essa non ne trova alcuna soddisfacente.

15.7 Critiche agli assiomi di Arrow

La dimostrazione di Arrow suscitò un vivace dibattito durato parecchi anni. In particolare, i critici cercarono di minare alla base la dimostrazione suggerendo che gli assiomi non fossero così naturali e ovvii quanto sembrano. Le critiche sono parzialmente giustificate nel dire che il modello di Arrow non corrisponde esattamente alla realtà, ma non riescono tuttavia a distruggere il nucleo fondamentale del teorema, cioè il fatto che nessun meccanismo di aggregazione delle preferenze riesce a soddisfare i requisiti di cui sarebbe desiderabile godere. Vediamo in maggiore dettaglio alcune critiche.

15.7.1 Critiche all'assioma di non banalità

L'assioma di non banalità è ovviamente poco criticabile: in pratica un gruppo include sempre più di un decisore e le alternative sono spesso più di due. Tuttavia, i sostenitori dei sistemi bipartitici hanno spesso sottolineato il fatto che essi corrispondano esattamente al caso lasciato scoperto dalla dimostrazione di Arrow. In effetti, secondo il *teorema di Black*⁴, se l'insieme delle alternative contiene solo due soluzioni x e y , si possono soddisfare tutti gli altri assiomi di Arrow con una costituzione che conta i decisori che preferiscono l'una e quelli che preferiscono l'altra e assegna la preferenza del gruppo alla soluzione con il maggior numero di sostenitori. Il teorema è noto con il nome di *teorema dell'elettore mediano* perché dimostra anche che, se è nota la forza della preferenza dei decisori per le due alternative, e se si distribuiscono gli elettori lungo una linea continua in proporzione alla forza di tale preferenza, l'alternativa vincente è quella preferita dal decisore mediano, cioè dal decisore che ha metà dei decisori da una parte e metà dall'altra.

Questo ha condotto gli estimatori del bipartitismo a considerarlo l'unica forma realizzabile di democrazia, e quindi a sostenere i sistemi elettorali che favoriscono l'aggregazione su due fronti alternativi. Questa posizione ha dei punti di forza, ma si può ovviamente obiettare che la riduzione delle soluzioni a due anche nei casi in cui non sono due richiede a sua volta dei procedimenti di scelta per i quali la teoria non offre garanzie di democraticità.

15.7.2 Critiche all'assioma del dominio universale

Critiche di natura teorica sono invece state avanzate all'assioma del dominio universale, cioè alla pretesa che una costituzione sia in grado di aggregare le preferenze qualunque siano le opinioni dei decisori. Sempre il teorema di Black mostra che, se le preferenze dei decisori rispettano alcune ipotesi restrittive, la votazione a maggioranza relativa (cioè il sistema di pluralità) rispetta tutti gli altri assiomi di Arrow. Questo significa che, se i decisori distribuiscono le proprie preferenze in modo opportuno, anziché avere qualsiasi profilo immaginabile, è possibile garantire una procedura democratica di aggregazione.

In particolare, la restrizione ipotizzata da Black richiede che le alternative siano ordinabili in una sequenza $X = (x_1, \dots, x_n)$ secondo un criterio qualsiasi (non importa che segua una logica di qualche tipo) e che (questo è il punto fondamentale) la relazione di preferenza di ogni decisore sia conforme a una funzione valore unimodale rispetto a tale sequenza. In parole poche, ogni decisore deve avere una soluzione preferita (o un insieme di soluzioni preferite consecutive) e poi deve considerare le soluzioni che precedono la preferita come sempre peggiori via via che si procede a ritroso lungo la sequenza, e le soluzioni che seguono la preferita come sempre peggiori via via che si procede in avanti lungo la sequenza. La Figura 15.7.2 illustra il risultato: i singoli decisori corrispondono alle curve unimodali (prima crescenti, poi decrescenti), che esprimono la forza della loro preferenza per le alternative (distribuite lungo l'asse delle ascisse). Gli assiomi di Arrow sono tutti rispettati, salvo quello del dominio universale, dal metodo di Condorcet, che porta alla vittoria della soluzione preferita dal decisore mediano, cioè da quello il cui massimo ha metà dei massimi associati agli altri decisori sulla sinistra e metà sulla destra.

Questa situazione ricorda chiaramente i periodi nei quali la vita politica di un paese si polarizza in una sequenza lineare di posizioni comprese fra due ali estreme, classicamente indicate come “sinistra” e “destra”. Se i decisori hanno una posizione politica preferita e considerano le altre sempre peggiori via via che ci si allontana

⁴Duncan Black (1908-1991), economista scozzese.

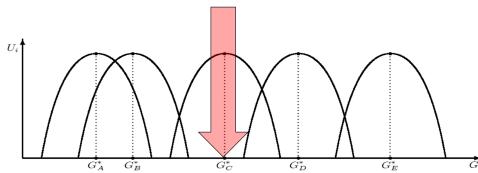


Figura 15.1: Cinque decisori esprimono le loro preferenze, tutte unimodali, per le alternative distribuite da sinistra verso destra lungo l'asse delle ascisse; il metodo di Condorcet porta a scegliere l'alternativa indicata dalla freccia, che è la preferita del decisore mediano, cioè di quello che ha due decisori sulla propria sinistra e due sulla destra.

in una o nell'altra direzione (anche in modo non simmetrico, purché monotono), il metodo di Condorcet consente di aggregare le preferenze soddisfacendo gli assiomi di Arrow. Se però alcuni decisori non rispettano questa *forma mentis* (per esempio, preferiscono le posizioni estreme, di destra o di sinistra che siano, a quelle intermedie, il sistema salta, e il teorema di Arrow torna ad esercitare il suo ruolo).

15.7.3 Critiche all'assioma della rilevanza binaria

Altre critiche sono state avanzate all'assioma di rilevanza binaria, secondo il quale la preferenza fra due soluzioni del gruppo deve dipendere solo dalle preferenze dei singoli fra quelle due soluzioni, e non da altre. Si è già osservato che questo equivale a imporre l'indipendenza dalle alternative irrilevanti, e serve a evitare che si possa manipolare la preferenza di gruppo cancellando alcune alternative o introducendone di nuove. Tuttavia, l'ipotesi di completa indipendenza sembra eccessivamente forte. Chiariamo il concetto con un esempio.

Esempio 90 Due individui si trovano al bar e decidono di ordinare la stessa bevanda. Per decidere quale, esprimono la propria preferenza ordinando le bevande disponibili. Il risultato è riportato nella tabella seguente.

	d_1	d_2
	Caffè	Tè
	Acqua	Caffè
	Limonata	Acqua
	Cola	Limonata
	Tè	Cola

Come negli altri esempi, si è supposto per semplicità che le preferenze dei decisori siano ordini totali. Dalla tabella è ragionevole dedurre che il caffè sia un buon candidato per una scelta comune, dato che è la bevanda preferita dal primo decisore e la seconda bevanda preferita dal secondo decisore. Tuttavia, l'assioma di rilevanza binaria richiede di considerare la preferenza di gruppo fra le due bevande come esattamente identica a quella che si avrebbe eliminando le altre, come riportato nella tabella seguente.

	d_1	d_2
	Caffè	Tè
	Tè	Caffè

Questa seconda situazione, però, appare del tutto diversa: le due bevande si contrappongono una all'altra in modo esattamente bilanciato, impedendo di fatto

una scelta (ognuna delle due scelte appare arbitrariamente a favore di uno dei due decisori).

Il punto fondamentale è che la presenza delle altre bevande consentiva in qualche modo non solo di esprimere l'ordine delle preferenze dei singoli, ma anche di misurarne la forza. Poteva anche essere un effetto ingannevole: magari il primo decisore gradiva più o meno ugualmente tutte le bevande, pur avendole ordinate in quel modo, e il secondo gradiva solo il tè, detestando tutte le altre. In realtà, per poter prendere una vera decisione occorrerebbe conoscere la forza della preferenza di ciascun decisore per ciascuna alternativa. Questo somiglia molto alla combinazione dei valori di molti attributi in una sola funzione di utilità, dove i decisori fanno la parte degli attributi. Se si ricordano le complicazioni introdotte da questo procedimento, si può ben intuire quanto poco esso sia applicabile all'aggregazione delle preferenze di molti decisori (potenzialmente centinaia di milioni, nel caso delle decisioni politiche di stati molto popolati) in una preferenza nazionale. Quindi la critica è valida, ma la contromisura è poco applicabile in pratica.

15.7.4 Critiche all'assioma di unanimità

Persino l'assioma di unanimità offre il fianco a critiche. Parrebbe ovvio che, se tutti i decisori hanno una preferenza, il gruppo nel suo complesso debba rispecchiarla fedelmente. Eppure, questo nega alcuni fondamenti sostenuti da diversi filoni del pensiero politico. Quando i cattolici parlano di *valori non negoziabili* e i liberali di *diritti fondamentali dell'uomo* intendono proprio che alcune preferenze (la libertà di parola rispetto alla censura, per esempio, o il diritto alla vita rispetto a quello alla scelta) non sono modificabili nemmeno se tutti i decisori la pensano al contrario.

D'altra parte, è anche chiaro che, se il fondamento di queste preferenze non è la preferenza dei singoli, si apre il problema di quale sia la sua fonte ultima, problema al quale sono state date diverse risposte che il modello di Arrow non prende in considerazione.

15.8 Impossibilità per i sistemi elettorali*

Esiste un teorema che richiama da vicino il teorema di Arrow, ma riguarda i sistemi elettorali, e precisamente l'assegnamento dei seggi ai partiti in una competizione elettorale. Come è noto, vi sono parecchi sistemi elettorali, da quello proporzionale puro ai sistemi maggioritari. Si può vedere un sistema elettorale come una regola di aggregazione e semplificazione delle preferenze degli elettori per i partiti esistenti. Le preferenze sono semplificate a priori, ipotizzando che ogni elettore scelga un solo partito. Si vuole ottenere una descrizione aggregata di tali preferenze che assegna ad ogni partito un opportuno numero di seggi in un parlamento.

Definizione 63 *Sia dato un insieme finito D di elettori e un insieme finito X di $n = |X|$ partiti. L'insieme dei vettori di voto è l'insieme $V = \{v \in \mathbb{N}^n : \sum_{x \in X} v_x = |D|\}$ dei vettori di n numeri interi con somma $|D|$.*

I vettori di voto contano il numero di elettori che preferiscono ciascun partito.

Definizione 64 *Dato un insieme S di seggi da assegnare in parlamento, si dice sistema elettorale qualsiasi funzione $s : V \rightarrow \mathbb{N}^n$ che trasforma qualsiasi vettore di voto $v \in V$ nel numero dei seggi assegnati a ciascun partito, in modo tale che $\sum_{x \in X} s_x(v) = |S|$.*

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

Questa definizione include i sistemi elettorali più vari, a partire da quelli a partito unico, che, per qualsiasi vettore di voto attribuiscono tutti i seggi al partito x_1 , passando per quelli che assegnano tutti i seggi al partito con la maggioranza relativa, fino a quelli che assegnano a ciascun partito un numero di seggi esattamente proporzionale al numero dei voti, con qualche meccanismo più o meno complicato di arrotondamento (questo punto è tutt'altro che banale).

Data questa definizione, ha senso indagare quali sistemi godano delle proprietà che in teoria un sistema elettorale dovrebbe garantire per il buon funzionamento di uno stato. Purtroppo, anche questo problema gode di un risultato di impossibilità analogo a quello di Arrow.

Teorema 27 *Se l'insieme dei partiti ha almeno tre elementi, nessun sistema elettorale gode delle seguenti proprietà:*

1. *remunerazione: i partiti senza voti non prendono seggi*

$$s_x(v) = 0 \text{ per ogni } v \in V \text{ tale che } v_x = 0$$

2. *monotonia: i partiti con più voti prendono almeno lo stesso numero di seggi (i voti sono molto più numerosi dei seggi, per cui passando dai voti ai seggi la grana si fa grossa)*

$$s_x(v) \geq s_y(v) \text{ per ogni } v \in V \text{ tale che } v_x > v_y$$

3. *maggioranza assoluta: un partito con la maggioranza assoluta dei voti ha la maggioranza assoluta dei seggi e viceversa*

$$v_x > \frac{|D|}{2} \Leftrightarrow s_x(v) > \frac{|S|}{2} \text{ per ogni } v \in V$$

4. *superadditività: la fusione di due partiti deve avere almeno lo stesso numero di seggi dei due partiti separati*

$$v = \begin{bmatrix} v_1 \\ \dots \\ v_x \\ \dots \\ v_y \\ \dots \\ v_n \end{bmatrix}, \quad v' = \begin{bmatrix} v_1 \\ \dots \\ v_x + v_y \\ \dots \\ 0 \\ \dots \\ v_n \end{bmatrix} \Rightarrow s_x(v) + s_y(v) \leq s_x(v')$$

La quarta proprietà merita qualche spiegazione: il vettore v' aggrega i voti dei partiti x e y in un solo partito (indicato come x per semplice comodità), mentre y rimane come semplice segnaposto nel vettore, senza voti. La proprietà di superadditività richiede che il nuovo partito abbia un numero di seggi non inferiore alla somma dei due partiti originari. Il partito fittizio non ha seggi, grazie alla proprietà di remunerazione. Questa proprietà ha lo scopo di evitare che un partito si frammenti in due per sfruttare l'incentivo offerto dal sistema elettorale, che consente di ottenere complessivamente più seggi scindendosi che aggregandosi.

Parte V

Modelli descrittivi

Capitolo 11

Modelli per i sistemi di trasporto

Ci occupiamo dei sistemi per il trasporto di persone, sia pubblici sia privati. Consideriamo *modelli descrittivi*, nei quali cioè si vuole *determinare il valore di alcune grandezze incognite a partire da quello di altre grandezze note*. Non consideriamo invece modelli decisionali, cioè non imponiamo noi il valore di alcuna grandezza in gioco. Il modello descrittivo verrà poi usato in supporto a modelli decisionali (aprire o chiudere strade, cambiare sensi unici, modificare l'ampiezza, istituire corsie riservate, modificare i cicli semaforici, introdurre rotonde, ecc...). Il modello

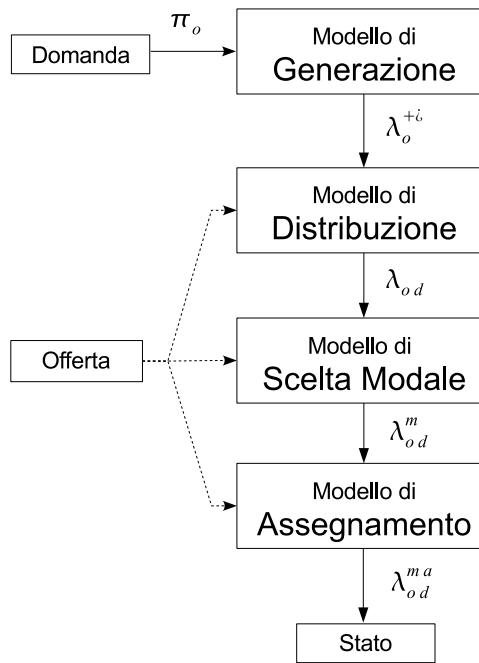


Figura 11.1: Sistemi di trasporto, modello a quattro stadi

che presentiamo è il più classico modello per i sistemi di trasporto. Viene detto *modello a quattro stadi* in quanto si compone di una catena di quattro sottomodelli, ognuno dei quali riceve come dati i risultati del precedente e fornisce come dati al successivo i propri risultati (Fig. 11.1). Questo ha due conseguenze immediate:

1. gli errori di modellazione compiuti in uno stadio si ripercuotono sui successivi, accumulandosi
2. ciascun sottomodello può essere realizzato in vari modi, dando origine a parecchi diversi modelli complessivi

In particolare, alcuni dati possono essere ricavati da misurazioni anziché dai modelli precedenti. Questo è possibile se il modello di trasporto viene usato per descrivere situazioni identiche o simili a quella attuale, per cui i valori misurati sono significativi, oltre ad essere esenti da errori di modellazione. D'altra parte, se il modello va usato per descrivere situazioni molto diverse dall'attuale, oppure se i dati misurati sono incompleti, se sono campionari e poco significativi o se sono affetti da errori di misurazione, l'uso di modelli offre una possibilità di sostituirli o correggerli.

Ogni sottomodello del modello a quattro stadi ha parametri che devono essere determinati attraverso una fase di *calibrazione*, nella quale si individuano i valori che avvicinano il più possibile i valori previsti dal modello a quelli registrati. La fase di calibrazione comporta l'uso di modelli decisionali (si tratta di decidere il valore dei parametri) in supporto a un modello descrittivo, che poi verrà usato in supporto a modelli decisionali, come si è detto. Tali modelli possono essere modelli di stima ai minimi quadrati, modelli di massima verosimiglianza, ecc...

Il modello a quattro stadi parte dalla descrizione di

- *domanda di trasporto*: popolazione e sue caratteristiche socio-economiche; distribuzione spaziale e temporale, abitudini, servizi e attrattori di traffico presenti sul territorio
- *offerta di trasporto*: rete stradale (distanze, caratteristiche di ampiezza e forma, cicli semaforici, caratteristiche degli incroci, ecc...), reti di trasporto pubblico (linee, fermate, distanze, orari, frequenze).

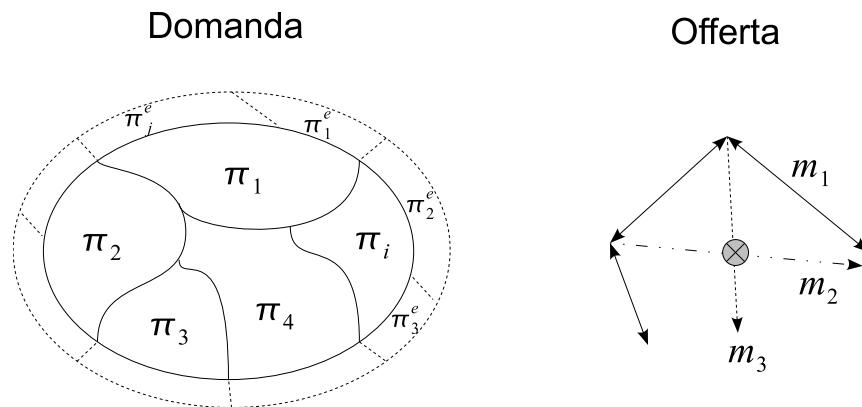


Figura 11.2: Sistemi di trasporto, modellazione di domanda ed offerta

Questi due elementi fondamentali sono spesso schematizzati come riportato in Fig. 11.2: per la domanda si utilizza un insieme di aree di origine π_i con una serie di aree periferiche π_j^e che esercitano un'influenza marginale sulla popolazione; per l'offerta si utilizza un grafo a più livelli dove ognuno rappresenta un differente mezzo di trasporto. In questa rappresentazione devono essere tenuti in considerazione sia i punti di giunzione fra diversi livelli (ad es. se è possibile scendere dal treno per prendere l'autobus), sia i vincoli derivanti dalla circolazione (semafori, sensi unici, ecc...).

In generale, l'offerta di trasporto deve rappresentare l'intera struttura viaria a disposizione nell'area considerata. È inoltre importante che la modellazione avvenga con un adeguato livello di dettaglio: questa rappresentazione verrà utilizzata in più livelli del modello.

Gli stadi componenti il modello sono i seguenti:

1. *generazione* o emissione della domanda (*chi fa il viaggio?*): definita una segmentazione della popolazione in fasce orarie, ed eventualmente in motivazioni di viaggio, categorie di viaggiatori, ecc... e una suddivisione spaziale della popolazione in aree dette *origini* (abbastanza grandi da essere gestibili, abbastanza piccole da rendere trascurabili gli spostamenti nella singola area), si procede a valutare il numero di spostamenti generati da ogni area

$$\lambda_o^+ = \phi_g(\pi_o, \dots)$$

2. *distribuzione* (*da dove a dove si viaggia?*): per ogni origine, si valuta la distribuzione dei viaggi fra le aree di arrivo (*destinazioni*). La segmentazione della popolazione serve a tener conto del fatto che per fasce orarie, motivazioni e categorie diverse di popolazione la distribuzione sarà anche diversa.

$$\lambda_{od} = \phi_d(\lambda_o^+, \dots)$$

3. *scelta modale* (*con che mezzo si viaggia?*): per ogni coppia origine-destinazione (O/D), si determina la distribuzione dei viaggiatori tra i diversi mezzi disponibili (privati e pubblici)

$$\lambda_{od}^m = \phi_m(\lambda_{od}, \dots)$$

4. *assegnamento* (*lungo quale percorso si viaggia?*): per ogni coppia O/D e per ogni mezzo, si determina il percorso effettivamente seguito sulla rete di trasporto

$$\lambda_{od}^{mp} = \phi_p(\lambda_{od}^m, \dots)$$

L'idea di fondo è esprimere il numero di spostamenti dalla zona o alla zona d con il mezzo m e lungo il percorso p (nella fascia oraria, per la motivazione e la categoria di persone scelte) come prodotto della popolazione presente nella zona o per la probabilità che ciascun individuo si sposti in d col mezzo m lungo il percorso p . Tale probabilità verrà espressa come prodotto di probabilità condizionate

$$\lambda_{od}^{mp} = \pi_o P(N=n) P(D=d|O=o) P(M=m|O=o, D=d) P(P=p|O=o, D=d, M=m)$$

Tramite questa catena di probabilità condizionate si tenta di modellare in modo trattabile il processo decisionale del singolo individuo. E' da notare come questa descrizione sia approssimata in quanto ipotizza che la catena sia orientata: la scelta della destinazione influenza la scelta del mezzo di trasporto, ma non viceversa. La stessa considerazione vale per tutte le altre scelte: ognuna influenza quelle a valle ma non è vero il contrario. Questa approssimazione rende quindi ogni livello di scelta indipendente dai successivi. L'ipotesi di fondo non descrive perfettamente la realtà ma, fra i possibili ordinamenti, la teoria ha scelto quello che riflette meglio le reali dinamiche di trasporto.

11.1 Modello di generazione

Si tratta di un modello molto semplice. Fissata la fascia oraria, ed eventualmente la motivazione e la categoria di viaggiatori di interesse, il modello stima il numero

di spostamenti generati da una certa zona come prodotto della popolazione per la probabilità che un individuo della zona (e della categoria specificata) compia uno spostamento all'interno della fascia oraria data. Più precisamente, ammettendo che un individuo possa compiere più spostamenti, si impiegano le probabilità che l'individuo compia un dato numero n di spostamenti, moltiplicato per il numero stesso:

$$\lambda_o^+ = \pi_o \sum_{n=1}^{n_{\max}} n P(N=n)$$

Si noti che la popolazione non va intesa come popolazione residente, ma come popolazione presente nell'area durante la fascia oraria di interesse (magari a causa di spostamenti precedenti). Quindi, le zone che generano traffico nelle fasce orarie mattutine, diventano zone di destinazione di traffico generato dalle zone lavorative nelle fasce orarie serali, quelle in cui la popolazione rientra da una giornata di lavoro. Nel corso della giornata le zone possono quindi assumere sia il ruolo di origini di traffico, sia di destinazioni.

Si possono stimare direttamente le probabilità $P(N=n)$, oppure esprimerele come funzioni (in genere lineari) di variabili $x_g^{(l)}$ che descrivano le caratteristiche socio-economiche della zona di origine, dalla categoria di viaggiatori, dalla fascia oraria e dalla motivazione di viaggio.

$$P(N=n) = \sum_l \beta_g^{(l)} x^{(l)}$$

dove gli *attributi* $x^{(l)}$ sono grandezze socio-economiche come note come il tipo di zona, il reddito, la diffusione dell'automobile, ecc... mentre i coefficienti $\beta_g^{(l)}$ dovranno essere calibrati rispetto ad un campione di dati significativo.

11.2 Modello di distribuzione

A questo punto, si tratta di stimare quanta parte dei viaggi generati si diriga verso ciascuna area di destinazione. L'idea alla base del modello di distribuzione è che ogni individuo sceglierà la propria destinazione come la più conveniente per soddisfare il bisogno all'origine del viaggio. Se i criteri che costituiscono l'utilità associata a un viaggio fossero perfettamente uniformi per tutti i viaggiatori di una certa zona di origine, essi si dirigerebbero alla zona che meglio li soddisfa. Se tali criteri fossero perfettamente noti, il loro comportamento sarebbe perfettamente prevedibile. Ad esempio, se l'unico criterio fosse la distanza (da minimizzare), tutti i λ_o^+ viaggi generati dalla zona o sarebbero diretti alla più vicina zona in grado di soddisfare la motivazione del viaggio.

Poiché così non è, senza abbandonare del tutto l'idea di un comportamento razionale e ottimizzante, si adottano *modelli di utilità aleatoria*, nei quali l'utilità del singolo individuo viene descritta come parzialmente nota e parzialmente aleatoria. Cercheremo quindi la probabilità che ogni individuo scelga una certa destinazione e valuteremo il flusso diretto a quella destinazione come prodotto del numero di individui in viaggio per tale probabilità.

$$\lambda_{od} = \lambda_o^+ P(D=d|O=o)$$

Le ipotesi di fondo sono:

1. il generico utente ha un insieme finito A di alternative disponibili; l'insieme è lo stesso per tutti (da qui l'esigenza di segmentare categorie di utenti diversi, per esempio dotati o no di patente di guida);

2. ogni decisore associa a ciascuna alternativa $a \in A$ un'utilità percepita U_a e sceglie l'alternativa che massimizza U_a
3. l'utilità U_a dipende da una serie di caratteristiche misurabili (*attributi*) $x_{(i)a}$ propri dell'alternativa e del decisore; quindi, il decisore sceglie un'alternativa confrontando gli attributi di quell'alternativa con quelli delle altre;
4. l'utilità U_a non è nota con certezza all'osservatore esterno, e pertanto deve essere rappresentata con una variabile aleatoria, pari alla somma di una componente nota (costituita da quegli attributi che l'osservatore può misurare) e di una componente aleatoria o *residuo* (costituita da quegli attributi che l'osservatore non può misurare):

$$U_a = V_a + \epsilon_a \quad a \in A$$

5. la dipendenza della componente deterministica V_a dagli attributi viene in genere modellata come lineare, con coefficienti uniformi per tutti gli individui. Gli errori compiuti con tale ipotesi vengono anch'essi attribuiti alla componente aleatoria.

$$V_a = \sum_i \beta_d^{(i)} x_a^{(i)} \quad a \in A$$

dove $x_a^{(i)}$ sono gli attributi dell'alternativa a e i coefficienti $\beta_d^{(i)}$ esprimono il peso dell'attributo i sull'utilità dell'individuo.

A questo punto, la probabilità di scegliere l'alternativa \bar{a} è

$$P(A = \bar{a}) = P(U_{\bar{a}} \geq U_a, \forall a \in A) = P(\epsilon_{\bar{a}} \geq \epsilon_a + V_a - V_{\bar{a}}, \forall a \in A)$$

Per valutare questa probabilità, occorre un'ipotesi sulla distribuzione stocastica dei residui. Il modello più diffuso in letteratura è il *modello Logit multinomiale*. Esso assume che i residui seguano *identicamente e indipendentemente una distribuzione di Gumbel a media nulla e varianza θ^2* .

$$F(x) = e^{-e^{-\frac{x}{\theta}} - \Phi}$$

dove Φ è la costante di Eulero.

Questa distribuzione è abbastanza simile alla distribuzione normale (quella tipica dei valori casuali), ma ha in più la proprietà di poter esprimere in forma chiusa la probabilità precedente.

$$P(A = \bar{a}) = \frac{e^{U_{\bar{a}}}}{\sum_{a \in A} e^{U_a}}$$

che ha un andamento sigmoidale piuttosto verosimile (Fig. 11.3). Qualora tutte le alternative avessero la stessa utilità deterministica ($U_a = \bar{U}, \forall a \in A$), il modello suggerisce che gli individui si disperdano uniformemente fra le diverse destinazioni ($P(A = a) = 1/|A|$).

Altre proprietà interessanti sono il fatto che le probabilità dipendono dalle differenze tra i valori delle utilità sistematiche, e non dal loro valore assoluto e quindi risultano indipendenti alla scelta dello zero per ciascun attributo. Inoltre, il modello risulta indipendente dalle alternative irrilevanti, cioè aggiungere o togliere alternative non altera l'ordinamento relativo fra quelle presenti: cambia il denominatore ma allo stesso modo per tutti i coefficienti, come se fossero moltiplicati per un identico fattore. Infine, esiste l'influenza della varianza dei residui: al crescere di θ , le

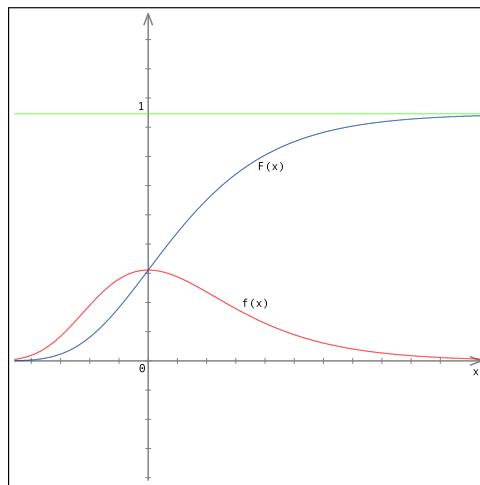


Figura 11.3: Distribuzione di Gumbel $F(x)$ e relativa funzione di densità di probabilità $f(x)$.

probabilità tendono ad avvicinarsi, cioè meno sono significativi i termini noti, meno si differenziano fra loro le scelte.

Il modello richiede una fase di calibrazione per valutare il parametro θ e i coefficienti $\beta_d^{(i)}$ dei diversi attributi.

Il modello Logit è ampiamente utilizzato anche nell'ambito del marketing, per descrivere il comportamento del consumatore posto di fronte ad un insieme di possibili prodotti o servizi da acquistare.

11.3 Modello di scelta modale

Stima quanta parte dei viaggi dall'origine o alla destinazione d utilizza il mezzo m . I modelli Logit multinomiali sono stati applicati per la prima volta a questa fase, estendendosi poi a quella di distribuzione, nella quale in precedenza si adottavano modelli diversi (ad esempio, il modello gravitazionale o quello a massima entropia).

Gli attributi tipici usati nel modello Logit per la scelta modale sono:

- tempo di viaggio
- costo monetario
- comodità
- tempo di avvicinamento al mezzo e di allontanamento dal mezzo
- tempo di attesa (nullo per il mezzo privato)

Trattandosi di criteri eterogenei, i coefficienti $\beta^{(i)}$ devono correlare fra loro le unità di misura, traducendo tutti gli attributi in termini di utilità adimensionale. Nel caso di costi (cioè escludendo solo la comodità), i coefficienti sono negativi.

Ricavare il peso $\beta_m^{(i)}$ assegnato ad ogni attributo $x_a^{(i)}$ dell'alternativa a richiede uno studio dedicato che può essere svolto secondo differenti modalità. Nel caso in cui si contatti un campione dell'utenza per rilevarne direttamente utilità e bisogni si

parla di metodi delle *preferenze dichiarate* o *stated preferences*. Sotto questa dicitura viene raccolto un insieme di metodi di pesatura (o *pricing*) dove il bacino d'utenza viene interpellato direttamente tramite opportuni questionari. Le domande sono strutturate in modo tale da ricavare, per ognuna delle alternative presenti, un valore monetario: si può chiedere all'intervistato quanto sarebbe disposto a pagare per evitare un peggioramento del servizio o, in alternativa, quanto sconto esigerebbe sul prezzo del servizio per sentirsi in un certo modo “risarcito” del calo di qualità subito. Un'ulteriore possibilità prevede di chiedere all'intervistato di determinare un tradeoff accettabile fra alternative differenti, dal quale poi è possibile stimare il valore assegnato ad ognuna. La seconda categoria di metodi, definita delle *preferenze rilevate* (*revealed preferences*), richiede l'osservazione del comportamento dell'utenza a fronte di modifiche della qualità o del prezzo delle alternative. I metodi appartenenti a questa categoria sono considerati i più affidabili dalle discipline *econometriche* in quanto si osservano i comportamenti reali dell'utenza, evitando le potenziali imprecisioni dovute alla incongruenze di valutazione che possono sorgere nel decisore messo di fronte ad un questionario. Nonostante questo difetto, i metodi delle preferenze dichiarate sono, per ovvi motivi, decisamente meno costosi ed impegnativi e permettono inoltre di assegnare un valore di utilità anche ad alternative che non sono ancora state implementate.

11.4 Modello di scelta del percorso

Questo modello parte dalle matrici O/D associate a ciascun mezzo e arriva a determinare il numero di individui che seguono ogni percorso. Nel caso dei mezzi pubblici, il modello è generalmente banale, dato che il percorso disponibile è uno solo, oppure si riduce ad un'ulteriore applicazione del modello Logit, oppure a un'applicazione di modelli statistici basati sulla frequenza delle corse (l'ipotesi è che l'individuo prenda il mezzo che passa per primo fra quelli che possono servire la sua richiesta).

Nel caso del trasporto privato invece, a partire dalle matrici O/D il modello assegna il valore del flusso di veicoli su ogni arco della rete stradale. In questo caso, la scelta del percorso è tutt'altro che banale perché molti percorsi sono disponibili e soprattutto perché il tempo di viaggio su ogni percorso dipende pesantemente dal numero di veicoli che lo seguono. I due aspetti sono quindi profondamente correlati, influendo l'uno sull'altro: un percorso più breve attira più traffico e diventa quindi automaticamente più lungo.

Quando un insieme di veicoli percorre una strada, il tempo che impiega è inversamente proporzionale alla sua velocità. La velocità di un veicolo è una variabile aleatoria, che il veicolo tipicamente tende a conservare costante. Se sulla strada ci sono però altri veicoli, la cosa comporta sorpassi, e questo è possibile solo se l'intervallo fra il veicolo precedente e quello prima ancora lo consente. Quindi, il tempo di viaggio medio tende a crescere al crescere del flusso. Sulle autostrade, si considera approssimativamente lineare la relazione fra tempo e flusso.

$$t_a = \alpha_a + \beta_a \frac{x_a}{K_a} \quad (11.1)$$

dove α_a , β_a e K_a sono parametri dell'arco di strada a . Ovviamente esiste un limite superiore al flusso sopportabile dal tratto prima dell'imbottigliamento dei viaggiatori come mostrato in Fig. 11.4. Per comprendere il significato del parametro K_a , si consideri la dinamica del flusso di veicoli in viaggio su di un tratto autostradale. Se un conducente che sta percorrendo la corsia di destra incontra un veicolo che viaggia ad una velocità inferiore, cercherà di spostarsi sulla corsia a sinistra per superarlo. Allo stesso modo, se un conducente che sta percorrendo la corsia di sorpasso

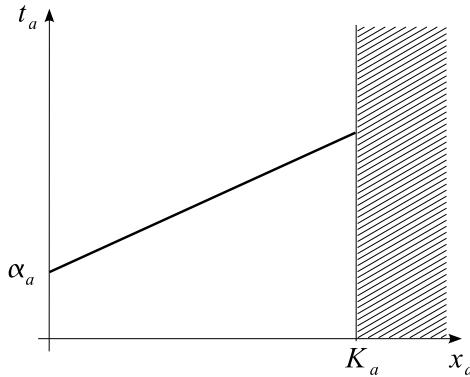


Figura 11.4: Tempo di percorrenza su arco autostradale

incontra un veicolo più lento, tenterà di indurlo a spostarsi sulla corsia di destra per mantenere la propria velocità. La capacità di spostarsi da una corsia all'altra è quindi una proprietà fondamentale per mantenere una velocità di crociera. Questa possibilità però non è sempre garantita. All'aumentare del traffico infatti, la facoltà di cambiare corsia diminuisce, poiché diventano più rari gli intervalli di spazio fra due veicoli entro i quali è possibile inserirsi con sicurezza, con conseguente diminuzione della velocità media del flusso veicolare. Quando il traffico è molto denso, i veicoli che si spostano sulla corsia di sorpasso si inseriscono in intervalli abbastanza stretti, costringendo chi li precede a decelerare, spesso bruscamente. Queste decelerazioni si propagano anche ai veicoli successivi, dando luogo a forti oscillazioni di velocità. In queste condizioni il flusso diventa instabile: si verificano rapide ed improvvise cadute di velocità, con possibile arresto della circolazione. Si definisce *capacità fisica* di una carreggiata autostradale quel valore del flusso veicolare al di là del quale il rischio della instabilità è considerato inaccettabile. Imponendo quindi che

$$x_a \leq K_a$$

la funzione lineare (11.1) diventa verosimile. Come si nota dalla Figura 11.4, l'andamento oltre K_a non è rappresentato in quanto non sarebbe significativo: oltre quel valore il flusso veicolare è in condizione di instabilità. Il valore di *capacità fisica* dipende dalla geometria della carreggiata, dalla composizione del traffico, dalle condizioni atmosferiche e dal comportamento dei conducenti.

Nelle strade urbane invece, anche quando non ci sono carreggiate separate, il sorpasso rimane difficoltoso o vietato. Questo vincolo imposto sulla circolazione urbana rende le strade di questo tipo simili ad una coda *fifo* in quanto tutti i veicoli che intraprendono il percorso devono necessariamente assestarsi alla velocità del più lento. Il tempo di percorrenza è descritto in modo semplificato dalla seguente relazione (vedi Fig. 11.5).

$$t_a = \bar{t}_a \left[1 + \alpha_a \left(\frac{x_a}{K_a} \right)^{\beta_a} \right] \quad (11.2)$$

dove \bar{t}_a rappresenta il tempo richiesto per la percorrenza del tratto quando sgombero (detto *tempo a vuoto*), K_a la *capacità fisica* della strada espressa in numero di veicoli, α_a e β_a sono parametri dipendenti dalle caratteristiche del tratto. Come si nota dal grafico in Figura 11.5, l'andamento del tempo di percorrenza t_a aumenta polinomialmente in funzione del flusso a differenza di quanto si aveva nel caso autostradale dove incrementava linearmente il tempo di percorrenza del tratto fino

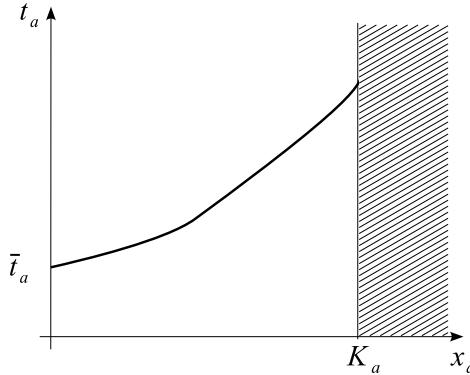


Figura 11.5: Tempo di percorrenza su arco stradale urbano

al raggiungimento del limite K_a . Anche in questo caso, la validità della (11.2) è garantita solo dove $x_a \leq K_a$.

Principi di Wardrop Per descrivere il comportamento del traffico privato si utilizzano i *Principi di Wardrop*. Essi tengono conto delle considerazioni fatte fino a questo punto e tentano di modellare il processo decisionale del singolo una volta messo di fronte al problema della scelta del percorso da seguire.

Siano P_{od} l'insieme dei percorsi da o a d , λ_{od} il flusso complessivo dall'origine o alla destinazione d ed il coefficiente binario δ_{pa} l'indicatore di appartenenza del tratto stradale a ad uno dei percorsi p utilizzati ($\delta_{pa} = 1$ in caso positivo, $\delta_{pa} = 0$ altrimenti). Sia inoltre la variabile x_a il flusso sul tratto stradale a e λ^p il flusso presente sul singolo percorso p . I vincoli di conservazione diventano:

$$\sum_{p \in P_{od}} \lambda^p = \lambda_{od} \quad \forall o, \forall d \quad (11.3)$$

$$x_a = \sum_p \lambda^p \delta_{pa} \quad \text{con} \quad \delta_{pa} = \begin{cases} 1 & \text{se } a \in p \\ 0 & \text{se } a \notin p \end{cases} \quad (11.4)$$

$$\lambda^p \geq 0 \quad \forall p \quad (11.5)$$

Primo principio di Wardrop *Il tempo di viaggio su ogni percorso usato è minore o uguale al tempo di viaggio che un singolo veicolo sperimenterebbe se usasse uno qualsiasi dei percorsi rimasti inutilizzati.* Ciò equivale a dire che ogni singolo utente cerca di minimizzare, in modo non cooperativo, il proprio tempo di viaggio. I flussi di traffico che soddisfano questo principio sono detti *di equilibrio*. Il punto di equilibrio è raggiunto quando nessun utente, tramite una decisione personale, può diminuire il proprio tempo di viaggio. Il primo principio di Wardrop afferma che:

$$T_{od}^* \leq T^p \quad \forall p \in P_{od} \quad (11.6)$$

$$\lambda^p (T^p - T_{od}^*) = 0 \quad \forall p \in P_{od} \quad (11.7)$$

In particolare, la (11.7) è detta *vincolo di complementarità* ed impone che, se il percorso p non è ottimo, il flusso su di esso è nullo e non passa alcun veicolo ($T^p > T_{od}^* \Rightarrow \lambda_p = 0$).

Secondo principio di Wardrop Si consideri ora il tempo totale T^p del percorso p dato dalla somma di tutti i tempi degli archi stradali a che lo compongono;

$$T^p = \sum_{a \in p} t_a(x_a)$$

Si consideri anche la seguente

$$T_{od}^* = \min_{p \in P_{od}} T^p$$

che rappresenta il tempo ottimo di percorrenza dall'origine o alla destinazione d . L'equazione (11.7) è la condizione necessaria di equilibrio ed assume lo stesso ruolo che le condizioni di Karush-Kuhn-Tucker ricoprono nell'ambito della Programmazione Matematica (vedere capitolo dedicato). Infatti, il secondo principio di Wardrop afferma che *all'equilibrio il tempo medio di viaggio è minimo*:

$$\varphi_x = \sum_{a \in A} \int_0^{x_a} t_a(x) dx \text{ è minimo}$$

Per dimostrare la relazione (11.7), si consideri la funzione φ si cerchi di minimizzarla nel rispetto dei vincoli (11.3), (11.4) e (11.5). Le condizioni di Karush-Kuhn-Tucker diventano:

$$l = \sum_{a \in A} \int_0^{x_a} t_a(x) dx + \sum_{od} \mu_{od} \left(\sum_{p \in P_{od}} \lambda^p - \lambda_{od} \right) + \sum_{a \in A} \mu_a \left(x_a - \sum_p \lambda^p \delta_{pa} \right) + \sum_p \mu_p \lambda^p$$

$$\frac{dl}{dx_a} = t_a(x_a) + \mu_a = 0 \implies \mu_a = -t_a(x_a)$$

$$\frac{dl}{d\lambda^p} = \mu_{od} - \sum_{a \in A} \mu_a \delta_{pa} + \mu_p = 0 \implies \sum_{a \in A} \delta_{pa} t_a(x_a) + \mu_{od} + \mu_p = 0$$

$$\mu_p = \mu_{od} - \sum_{a \in A} t_a(x_a) \delta_{pa} \geq 0 \implies T^p - T^* \geq 0$$

$$x_a = \sum_p \lambda^p \delta_{pa}$$

$$\lambda_{od} = \sum_{p \in P} \lambda^p$$

$$\mu_p \lambda^p = 0$$

$$\mu_p \geq 0$$

dalle quali si ottiene la (11.7). Quindi si minimizza il tempo *medio* di viaggio nell'intervallo $(0; x_a)$ ma sulla rete *tutti* subiscono il tempo associato al flusso massimo x_a che è maggiore. Come si può notare quindi, la situazione di equilibrio che si configura non è ottimale per alcun viaggiatore.

Il paradosso di Braess Un esempio significativo delle conseguenze dei principi di Wardrop è costituito dalla seguente situazione. Si consideri una rete stradale, modellata nel grafo di Fig. 11.6, dove l'andamento del tempo di percorrenza dei

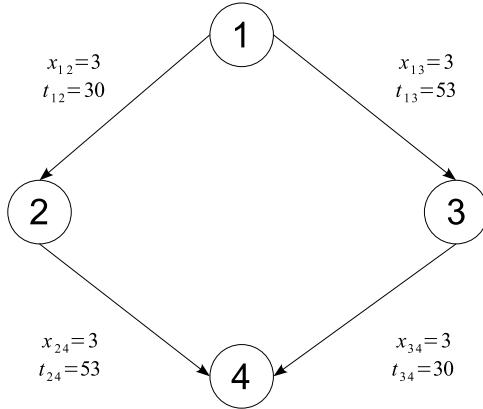


Figura 11.6: Paradosso di Braess, situazione di equilibrio iniziale

singoli tratti è descritto dalle seguenti equazioni

$$\begin{aligned} t_{12} &= 10x_{12} \\ t_{13} &= x_{13} + 50 \\ t_{24} &= 10x_{24} \\ t_{34} &= x_{34} + 50 \end{aligned}$$

e la rete soggetta ad una domanda di trasporto pari a $\lambda_{14} = 6$. Seguendo il primo principio di Wardrop, la situazione di ottimalità si raggiunge suddividendo egualmente il flusso in uscita dal nodo 1, ottenendo così tempi di percorrenza equivalenti su entrambi i percorsi possibili:

$$\begin{array}{ll} p_1 = \{1, 2, 4\} & t_1 = 83 \\ p_2 = \{1, 3, 4\} & t_2 = 83 \end{array}$$

Il decisore si pone però il problema di diminuire il tempo di percorrenza, migliorando così la rete stradale sotto il suo controllo. Per raggiungere in modo semplice questo obiettivo egli decide di aggiungere un nuovo tratto molto veloce che collega i nodi 2 e 3: il tempo di percorrenza del nuovo passaggio è descritto dalla seguente

$$t_{23} = x_{23} + 10$$

e la situazione che si viene a creare è quella rappresentata in Fig. 11.7. Considerando i percorsi possibili nella situazione finale, si ha:

$$\begin{array}{ll} p_1 = \{1, 2, 4\} & t_1 = 92 \\ p_2 = \{1, 3, 4\} & t_2 = 92 \\ p_3 = \{1, 2, 3, 4\} & t_3 = 92 \end{array}$$

Il primo principio di Wardrop può essere interpretato anche dal punto di vista singolare: ognuno tenta di minimizzare il tempo di percorrenza del proprio percorso in modo assolutamente egoistico, senza nessun tipo di coordinazione, portando ad una situazione finale che può essere peggiore per tutti. Il paradosso consiste proprio nell'aspetto di peggioramento per ognuno: se anche uno solo traesse beneficio dalla situazione finale, si trattrebbe di prepotenza di chi trova giovamento nei confronti di chi invece sarebbe costretto ad allungare il proprio viaggio. Si nota come nella situazione finale con la nuova strada ci sia una parte di viaggiatori (pari ad una

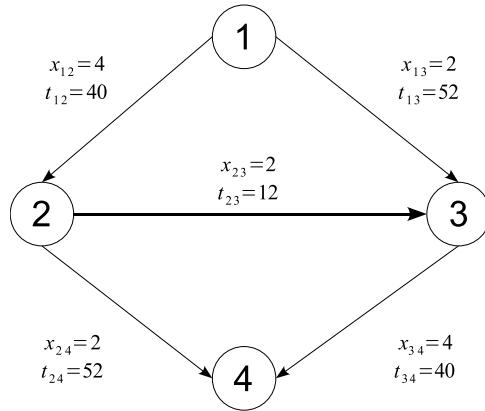


Figura 11.7: Paradosso di Braess, situazione di equilibrio finale

sola unità di flusso) che sceglie di cambiare percorso, passando dal nodo 2 per poter usufruire del nuovo tratto che si sa essere molto rapido. Questa decisione porta ad una perdita di flusso da parte delle strade più lente, che diventano leggermente più veloci, in favore di quelle più scorrevoli che diventano decisamente più lente. In questo caso, la situazione finale vede un deciso aumento del tempo di percorrenza di *tutti* i viaggiatori che da 83 passa a 92. La nuova situazione di equilibrio porta quindi a percorsi più lenti: aprire una nuova strada ha portato ad un netto peggioramento. Questi sono fenomeni che si osservano nella realtà e questo paradosso risulta significativo anche nell'ambito della Teoria dei Giochi: in questo caso tutti i giocatori (i viaggiatori) defezionano, tentati da una strada veloce, portando ad un'inevitabile punizione collettiva. L'aspetto di non-collaborazione del primo principio di Wardrop è così verificato; nel caso in cui tutti collaborassero, l'apertura di una nuova strada non modificherebbe i flussi in quanto tutti i viaggiatori la ignorerebbero (scegliendo di non defezionare ma di collaborare): basta che anche uno solo decida di farsi tentare per portare inevitabilmente alla defezione di massa.

Capitolo 12

Modelli di teoria delle code

Capitolo 13

Modelli di simulazione a eventi discreti

Capitolo 14

Modelli dinamici

Appendice A

Richiami di Analisi Matematica

Queste note riassumono alcuni concetti di Analisi Matematica necessari per affrontare i fondamenti della Programmazione Matematica (Cap. 1).

A.1 Richiami sulle funzioni di una sola variabile

Una funzione di una sola variabile $u = f(x)$ è una legge che fa corrispondere a ogni valore x di un opportuno insieme S in \mathbb{R} uno o più valori reali $u \in \mathbb{R}$.

$$u = f(x) \quad \text{oppure}$$

In generale, si intende che $f(x)$ assuma uno e un solo valore per ogni x in S .

Al variare di P in S , il punto $(P, f(P))$ descrive un luogo geometrico che si chiama *grafico* della funzione, ed è in generale una linea sul piano x, f . La Programmazione Matematica si occupa di determinare massimi e minimi di tale linea, con i relativi punti di massimo e minimo.

Teorema di Lagrange (teorema del valor medio) Se $f(x)$ è continua in un intervallo $[a; b]$ e derivabile in $(a; b)$, esiste un punto $\xi \in (a, b)$ tale che $f(b) - f(a) = (b - a) f'(\xi)$.

Questo teorema consente di esprimere il valore di una funzione nell'intorno di un punto x^* . Infatti, per ogni coppia di punti x^* e x in $[a; b]$ risulta

$$f(x) = f(x^*) + (x - x^*) f'(\xi)$$

dove ξ è un opportuno punto compreso fra x^* e x .

Formula di Taylor e formula di MacLaurin Se $f(x)$ è continua in un intervallo $[a; b]$ con le sue derivate sino all'ordine $n - 1$ ed esiste la sua derivata n -esima in $(a; b)$, per ogni coppia di punti x^* e x in $[a; b]$ risulta

$$\begin{aligned} f(x) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(x^*)}{k!} (x - x^*)^k + \frac{f^{(n)}(\xi)}{n!} (x - x^*)^n = \\ &= f(x^*) + \frac{f'(x^*)}{1!} (x - x^*) + \dots + \frac{f^{(n-1)}(x^*)}{(n-1)!} (x - x^*)^{n-1} + \frac{f^{(n)}(\xi)}{n!} (x - x^*)^n \end{aligned}$$

dove $\frac{f^{(n)}(\xi)}{n!} (x - x^*)^n$ si chiama *resto nella forma di Lagrange*.

Se in x^* esiste anche la derivata n -esima, si può scrivere

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x^*)}{k!} (x - x^*)^k + \frac{(x - x^*)^n}{n!} \left(f^{(n)}(x^*) + \alpha_n(x) \right) =$$

dove $\lim_{x \rightarrow x^*} \alpha_n(x) = 0$, oppure

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x^*)}{k!} (x - x^*)^k + o_n(x)$$

dove $o_n(x) = \frac{\alpha_n(x)}{n!} (x - x^*)^n$ si chiama *resto nella forma di Peano* ed è *infinitesimo di ordine superiore a n* rispetto a $(x - x^*)$, cioè $\lim_{x \rightarrow x^*} \frac{o_n(x)}{(x - x^*)^n} = 0$.

Il vantaggio della forma di Peano consiste nel far dipendere tutti i coefficienti dalle derivate di f in x^* , lasciando un termine infinitesimo che si può spesso eliminare attraverso un passaggio al limite.

Per $x^* = 0$, la formula di Taylor dà luogo alla *formula di MacLaurin*.

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(0)}{k!} x^k + o_n(x) \quad \text{con} \quad \lim_{x \rightarrow 0} \frac{o_n(x)}{x^n} = 0$$

A.2 Funzioni di più variabili

Una funzione di più variabili $u = f(P)$ è una legge che fa corrispondere a ogni punto P di un opportuno insieme S in \mathbb{R}^n uno o più valori reali $u \in \mathbb{R}$.

$$u = f(P) \quad \text{oppure} \quad u = f(x_1, x_2, \dots, x_n)$$

In generale, si intende che $f(P)$ assuma uno e un solo valore per ogni P in S .

Al variare di P in S , il punto $(P, f(P))$ descrive un luogo geometrico che si chiama *grafico* della funzione, ed è in generale una (iper)superficie a n dimensioni in uno spazio a $n + 1$ dimensioni. La Programmazione Non Lineare si occupa di determinare massimi e minimi di tale superficie, con i relativi punti di massimo e minimo.

Limiti di funzioni a più variabili Il concetto di limite si estende banalmente da funzioni di una a funzioni di più variabili.

$$\lim_{P \rightarrow P^*} f(P) = l$$

significa che per ogni $\epsilon > 0$ esiste un valore $\delta_\epsilon > 0$ tale che $|f(P) - l| < \epsilon$ per ogni $P \in \mathcal{U}_{P^*, \delta_\epsilon}$, cioè per ogni P tale che $\|P - P^*\| < \delta_\epsilon$.

Geometricamente, ciò significa che per ogni intervallo di semiampiezza ϵ intorno a l si riesce a trovare una sfera di raggio δ_ϵ centrata in x^* tale che il valore di f in tutti i punti della sfera cade nell'intervallo.

La distanza $\|P - P^*\|$ va definita opportunamente, ed è in genere quella euclidea.

$$\|P - P^*\| = \sqrt{(x_1 - x_1^*)^2 + (x_2 - x_2^*)^2 + \dots}$$

Una funzione $f(P)$ è continua in P^* quando

$$\lim_{P \rightarrow P^*} f(P) = f(P^*)$$

Composizione di funzioni Supponiamo che $z = f(P) = f(x_1, \dots, x_n)$ sia definita in un sottoinsieme S del piano \mathbb{R}^n e che le variabili x_i siano a loro volta funzioni $x_i = \phi_i(Q) = \phi_i(\xi_1, \dots, \xi_m)$ in un sottoinsieme Σ di \mathbb{R}^m . Allora z risulta essere funzione composta delle variabili ξ_j in Σ .

Si dimostra che, se ciascuna funzione ϕ_i ammette limite finito l_i per $Q \rightarrow Q_0$ e se f è continua in $P^* = (l_1, \dots, l_n)$, allora la funzione composta è continua in Q_0 e il limite è $f(\phi(Q_0), \psi(Q_0))$. Ad esempio, per $n = 2$:

$$\begin{cases} z = f(x_1, x_2) \\ x_1 = \phi_1(Q) \\ x_2 = \phi_2(Q) \\ \lim_{Q \rightarrow Q_0} \phi_1(Q) = l_1 \\ \lim_{Q \rightarrow Q_0} \phi_2(Q) = l_2 \\ f(P) \text{ continua in } P^* = (l_1, l_2) \end{cases} \Rightarrow \begin{cases} z = f(\phi_1(Q), \phi_2(Q)) \text{ è continua in } Q_0 \\ \lim_{Q \rightarrow Q_0} f(\phi_1(Q), \phi_2(Q)) = f(\phi_1(Q_0), \phi_2(Q_0)) \end{cases}$$

Derivate parziali Data una funzione di due variabili x e y , definita in un campo (insieme aperto e non vuoto), si attribuisca alla sola x un incremento $h \neq 0$, mantenendo invariata la y . Con h piccolo, il nuovo punto è ancora valido per la funzione. Si può definire il *rappporto incrementale parziale* e il suo limite, che è la *derivata parziale* rispetto a x . Analoghe sono le definizioni per la variabile y .

$$f_x = \frac{\partial f}{\partial x} = \lim_{h \rightarrow 0} \frac{f(x+h, y) - f(x, y)}{h}$$

Esempio: $z = e^{x/y}$ ha derivate parziali $z_x = \frac{1}{y} e^{x/y}$ e $z_y = -\frac{x}{y^2} e^{x/y}$

Le derivate parziali ammettono un'interpretazione geometrica: sono le tangenti trigonometriche degli angoli che le rette tangenti alla superficie $(P, f(P))$ in P e parallele ai piani xz e yx formano con gli assi x e y .

Esempio: $z = x^2 + 2y^2$ è un paraboloido con vertice nell'origine. Si vede che le tangenti vanno via via crescendo quando ci si allontana dall'origine e che crescono più in fretta rispetto alla y che rispetto alla x .

Una funzione può essere discontinua, ma derivabile parzialmente, al contrario di quanto avviene per le funzioni di una sola variabile. Basta infatti che siano continte le sue restrizioni agli assi.

Esempio:

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{per } (x, y) \neq (0, 0) \\ 0 & \text{per } (x, y) = (0, 0) \end{cases}$$

In $(0, 0)$ ammette entrambe le derivate parziali, perché lungo gli assi è identicamente nulla. D'altra parte, non esiste il limite per $(x, y) \rightarrow (0, 0)$; infatti, se ci si avvicina lungo la bisettrice $y = x$, f ha limite $1/2$; se ci si avvicina lungo l'altra bisettrice $y = -x$ ha limite $-1/2$; se ci si avvicina lungo gli assi, ha limite 0 .

Derivazione delle funzioni composte di una e più variabili Sia $f(x, y)$ definita in un campo A , con x e y entrambe funzioni di t in un intervallo $(a; b)$ e supponiamo che (x, y) resti in A mentre t varia in $(a; b)$. Allora, f è funzione composta di t nell'intervallo $(a; b)$.

Se $f \in C^1(A)$ e ϕ e ψ sono in $C^1(a, b)$, allora la funzione composta è $C^1(a, b)$ e la sua derivata è

$$f'(t) = f_x(\phi(t), \psi(t))\phi'(t) + f_y(\phi(t), \psi(t))\psi'(t)$$

Dimostrazione: La dimostrazione si basa sul teorema di Lagrange: per piccoli incrementi di t , x e y cambiano poco e i punti $(x + \delta x, y)$ e $(x + \delta x, y + \delta y)$ cadono in A . Quindi la variazione δz è la somma dell'incremento dovuto alla variazione di x e di quello dovuto alla variazione di y .

$$\begin{aligned} f(t+h) - f(t) &= f(\phi(t+h), \psi(t+h)) - f(\phi(t), \psi(t)) = \\ &= f(\phi(t+h), \psi(t+h)) - f(\phi(t+h), \psi(t)) + f(\phi(t+h), \psi(t)) - f(\phi(t), \psi(t)) \\ &= f_y(\phi(t+h), \eta)(\psi(t+h) - \psi(t)) + f_x(\xi, \psi(t))(\phi(t+h) - \phi(t)) \end{aligned}$$

dove $\eta \in (t, t+h)$ e $\xi \in (t, t+h)$. Quindi, tendendo al limite per $h \rightarrow 0$:

$$\begin{aligned} f'(t) &= \lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h} = \\ &= \lim_{h \rightarrow 0} \left[f_y(\phi(t+h), \eta) \frac{\psi(t+h) - \psi(t)}{h} + f_x(\xi, \psi(t)) \frac{\phi(t+h) - \phi(t)}{h} \right] \end{aligned}$$

per la continuità di f_x e f_y e per la definizione di derivata, segue la tesi.

Analogamente, se le n variabili x_i sono funzioni di m variabili ξ_j

$$f_{\xi_j} = \frac{\partial f}{\partial \xi_j} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \xi_j}$$

Basta ripetere il ragionamento, tenendo ogni volta fisse $m - 1$ variabili.

Derivata direzionale Consideriamo una funzione f definita in un campo A , un punto $P^* \in A$ e un asse r , con coordinate misurate a partire da P^* . Se consideriamo i valori di f solo in corrispondenza ai punti dell'asse r , si possono definire un rapporto incrementale e una derivata rispetto a r . Questa è la tangente trigonometrica dell'angolo che la retta tangente alla superficie in P e parallela al piano rz forma con l'asse r .

Abbiamo quindi infinite possibili derivate: le derivate parziali sono quelle per cui r coincide con uno degli assi. Fortunatamente, per conoscere le infinite derivate direzionali, è sufficiente conoscere le n derivate parziali. Scrivendo x e y come funzioni di r con le coordinate polari:

$$x = x^* + r \cos \theta \text{ e } y = y^* + r \sin \theta$$

la funzione f ristretta ai punti dell'asse r si può valutare come funzione composta, e la derivata direzionale si può ottenere come sopra descritto, ottenendo

$$f_r = f_x \frac{dx}{dr} + f_y \frac{dy}{dr} = f_x \cos \theta + f_y \sin \theta$$

Gradiente Il vettore le cui componenti sono le derivate parziali è detto *gradiente* della superficie. Individua un *campo vettoriale*, che sarebbe una funzione di più variabili a valori vettoriali.

$$\nabla f = \begin{bmatrix} f_x \\ f_y \end{bmatrix}$$

L'asse r è univocamente individuato dal suo *versore* u_r (vettore di norma unitaria)

$$u_r = \begin{bmatrix} \cos \theta_1 \\ \dots \\ \cos \theta_n \end{bmatrix}$$

dove θ_i è l'angolo che r forma con l'asse x_i e i coseni si chiamano *coseni direttori*.

Quindi, la derivata direzionale lungo l'asse r è pari al prodotto scalare del vettore gradiente con il versore di r .

$$f_r = f_x \cos \theta + f_y \sin \theta = (\nabla f)^T \cdot u_r$$

e questo implica che *la derivata direzionale è massima quando viene calcolata nella direzione del gradiente* e nulla in direzione ortogonale.

Il gradiente indica la direzione di più rapido incremento della funzione. Dette *linee di livello* i luoghi dei punti in cui $f(P) = k$, cioè la funzione ha valore uniforme, *il gradiente è in ogni punto normale alle linee di livello*.

Piano tangente La relazione fra le derivate direzionali si traduce in una relazione geometrica fra le rette tangenti al grafico di f . Infatti, i punti della tangente lungo l'asse r in $(x^*, y^*, f(x^*, y^*))$ si possono scrivere come

$$z - f(x^*, y^*) = f_r(0)r = [f_x(x^*, y^*) \cos \theta + f_y(x^*, y^*) \sin \theta]r$$

ma poiché $r \cos \theta = x - x^*$ e $r \sin \theta = y - y^*$:

$$z - f(x^*, y^*) = f_x(x^*, y^*)(x - x^*) + f_y(x^*, y^*)(y - y^*)$$

che rappresenta un piano: *tutte le rette tangenti a una superficie regolare in un punto giacciono su un piano*. Questo piano si dice *piano tangente*.

Formula di Taylor per funzioni di più variabili Ora possiamo tirare le somme: dato un punto P^* , approssimiamo il valore di f in qualsiasi punto P nell'intorno di P^* semplicemente tracciando un asse da P^* a P , considerando la restrizione di f a tale asse ed approssimandola con la formula di Taylor. In essa compare la derivata direzionale, che però si può esprimere in funzione delle derivate parziali.

Se $f(x, y) \in C^n(A)$, si consideri un punto $P^* = (x^*, y^*)$ e un punto $P = (x, y)$ tale che il segmento P^*P cada interamente in A . I punti del segmento hanno coordinate $(x^* + t(x - x^*), y^* + t(y - y^*))$ con $t \in [0; 1]$. Lungo il segmento, f è funzione composta di t . La formula di MacLaurin fornisce il valore di f lungo il segmento (in particolare in $t = 1$):

$$f(1) = \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!} 1^k + \frac{f^{(n)}(\xi)}{n!} 1^n$$

con un opportuno $\xi \in (0; 1)$. Arrestandosi al primo termine ($n = 1$), considerando $t = 1$ e riscrivendo nelle variabili x e y , si ottiene

$$f(x, y) = f(x^*, y^*) + [f_x(P(\xi))(x - x^*) + f_y(P(\xi))(y - y^*)]$$

dove $P(\xi) = (x^* + \xi(x - x^*), y^* + \xi(y - y^*))$.

Se le derivate prime sono continue, al resto in forma di Lagrange si può sostituire il resto in forma di Peano

$$f(x, y) = f(x^*, y^*) + [f_x(x^*, y^*)(x - x^*) + f_y(x^*, y^*)(y - y^*)] + o_1(x, y)$$

con $\lim_{P \rightarrow P^*} \frac{o_1(P)}{\|P - P^*\|} = 0$.

A.3 Linee in forma parametrica

Consideriamo due funzioni reali di variabile reale t definite e continue in un *insieme base* S . Se le interpretiamo come coordinate cartesiane di un punto, il loro insieme definisce una *linea in forma parametrica*. Il parametro t stabilisce un orientamento sulla linea.

Se un punto $P(t) = (x(t), y(t))$ di una linea in forma parametrica deriva da diversi valori di t , cioè se la linea si ripiega su sé stessa, si parla di *punto multiplo*.

Esempio:

$$\begin{cases} x(t) = t^2 \\ y(t) = t^3 - t \end{cases} \quad \text{con } t \in [-2, 2]$$

Esempio:

$$\begin{cases} x(t) = t^2 \\ y(t) = t^2 \end{cases} \quad \text{con } t \in [-1, 1]$$

Una *linea semplice* è priva di punti multipli. Essa crea una corrispondenza biunivoca fra punti e valori di t .

Vettore tangente Si può estendere al vettore $P(t)$ il concetto di derivata, sempre basandolo sul rapporto incrementale: il risultato è un vettore (detto *vettore tangente* alla linea) le cui componenti sono le derivate delle singole componenti.

$$P'(t) = \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix}$$

Il piano normale al vettore tangente si dice *piano normale* alla linea.

Linee regolari Il concetto di linea in forma parametrica include un insieme di oggetti più ampio di quello suggerito dall'intuizione. Ad esempio, $x(t) = 1$ e $y(t) = 2$ è una linea in forma parametrica, pur riducendosi geometricamente a un solo punto. Ancora, la curva di Peano è una linea in forma parametrica, anche se geometricamente ricopre un intero quadrato.

Per evitare queste patologie, ci concentriamo sulle *linee regolari*:

1. *semplici*
2. aventi come *insieme base* un intervallo (a, b)
3. dotate di *vettore tangente continuo e non nullo* (cioè le singole derivate sono continue e non contemporaneamente nulle)

Esempio:

$$\begin{cases} x(t) = \sin t \\ y(t) = \cos t \end{cases} \quad \text{con } t \in [-\pi/2, \pi/2] \text{ è una linea regolare}$$

Esempio:

$$\begin{cases} x(t) = \sin t \\ y(t) = \cos t \end{cases} \quad \text{con } t \in [-\pi, \pi]$$

è un caso limite (*ciclo*) perché il punto $P(-\pi) = P(\pi) = (0, -1)$ è doppio.

Esempio:

$$\begin{cases} x(t) = t^2 \\ y(t) = t^3 \end{cases} \quad \text{con } t \in [-1, 1]$$

è semplice, ma non regolare: perché il vettore tangente $P'(t) = [2t \ 3t^2]^T$ si annulla in $(0, 0)$.

Rappresentazioni Una linea orientata ha infinite rappresentazioni parametriche: basta sostituire t con $t = \phi(\tau)$, dove ϕ è una funzione strettamente monotona in un intervallo. Per mantenere la regolarità, occorre e basta che ϕ sia continua e abbia derivata continua e mai nulla.

In particolare, la *rappresentazione cartesiana* impiega come parametro una delle coordinate. Una linea regolare ammette sempre una rappresentazione cartesiana locale, cioè valida nell'intorno di un punto $P^* = P(t_0)$: basta invertire una delle funzioni componenti, dato che non possono avere tutte derivata nulla. Però la rappresentazione cartesiana in generale copre solo un intorno di P^* .

Esempio (elica cilindrica):

$$\begin{cases} x(t) = R \cos t \\ y(t) = R \sin t \\ z(t) = \frac{p}{2\pi}t \end{cases} \quad \text{con } t \in \mathbb{R}$$

invertendo la terza relazione $z(t) = \frac{p}{2\pi}t \Leftrightarrow t(z) = \frac{2\pi}{p}z$, si traduce in

$$\begin{cases} x(t) = R \cos \frac{2\pi z}{p} \\ y(t) = R \sin \frac{2\pi z}{p} \end{cases} \quad \text{con } z \in \mathbb{R}$$

Altre rappresentazioni cartesiane si ottengono invertendo una delle prime due relazioni, ma esse valgono solo per un semigiro dell'elica, cioè in un intorno del punto scelto.

Appendice B

Richiami di Algebra Lineare

B.1 Combinazioni di vettori

Definizione 65 *Dato un insieme di m vettori g_j nello spazio \mathbb{R}^n , si dice combinazione lineare di tali vettori qualsiasi vettore ottenuto moltiplicando i primi per coefficienti reali e sommando i risultati:*

$$f = \sum_{j=1}^m \mu_j g_j$$

Una combinazione lineare si dice:

- combinazione conica quando i coefficienti μ_j sono non negativi:

$$\mu_j \geq 0 \text{ per } j = 1, \dots, m$$

- combinazione affine quando i coefficienti μ_j hanno somma unitaria:

$$\sum_{j=1}^m \mu_j = 1$$

- combinazione convessa quando è sia conica sia affine:

$$\mu_j \geq 0 \text{ per } j = 1, \dots, m \text{ e } \sum_{j=1}^m \mu_j = 1$$

Appendice C

Richiami di calcolo delle probabilità

Seguiremo un approccio assiomatico, cioè definiremo le proprietà matematiche degli oggetti che ci occorrono, senza entrare nei (grossi) problemi filosofici di come dare loro un significato pratico. Vi sono approcci molto diversi alla soluzione di tali problemi.

1. È dato uno *spazio degli eventi* Ω , che contiene tutti i possibili *esiti* ω di un esperimento casuale; è anche data una collezione \mathcal{A} di sottoinsiemi dello spazio Ω detti *eventi*. Altrove nel corso gli esiti sono definiti scenari.
2. Ad ogni evento $A \in \mathcal{A}$ è associato un numero reale non negativo $P(A)$, detto *probabilità* di A

$$P : \mathcal{A} \rightarrow \mathbb{R}^+$$

3. L'intero spazio è un evento (*evento certo*) di probabilità pari a 1

$$P(\Omega) = 1$$

4. Se due eventi A e B contengono esiti distinti, la probabilità della loro unione coincide con la somma delle loro probabilità

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

5. Se A_n è una successione di eventi contenuti ciascuno nel precedente e al tendere di n all'infinito la loro intersezione tende all'insieme vuoto, allora $P(A_n)$ tende a zero:

$$A_n \rightarrow \emptyset \Rightarrow \lim_{n \rightarrow \infty} P(A_n) = 0$$

Dagli assiomi derivano le seguenti conseguenze o teoremi.

Teorema della probabilità totale La probabilità dell'unione di due o più eventi, cioè la probabilità che se ne verifichi almeno uno, è la somma delle probabilità dei singoli eventi meno la somma delle probabilità delle intersezioni due a due, più la somma delle probabilità delle intersezioni a tre a tre e così via. Ad esempio:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Se gli eventi sono a due a due incompatibili, è la semplice somma delle probabilità.

Probabilità condizionata La *probabilità condizionata* di A dato B , indicata con $P(A|B)$ è la probabilità che l'evento A ha di verificarsi quando si sappia che B si è verificato:

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

Teorema della probabilità composta La probabilità dell'intersezione di due o più eventi, ovvero la probabilità che essi si verifichino tutti è $P(A \cap B) = P(A|B)P(B)$.

Se la probabilità di A dato B , $P(A|B)$, è uguale a $P(A)$, i due eventi vengono definiti *indipendenti stocasticamente* (o probabilisticamente) e $P(A \cap B) = P(A)P(B)$.

Teorema di Bayes Dato un insieme finito o numerabile di eventi A_i , a due a due incompatibili, quando si sappia che si è verificato un evento E e che se E si verifica, allora si verifica necessariamente uno degli eventi di tale insieme (ed uno solo, dato che sono incompatibili), la probabilità a posteriori di A_i si può calcolare da quelle a priori degli eventi A_i e da quelle condizionate $P(E|A_i)$:

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_j P(E|A_j)P(A_j)}$$

In altre parole, se si conoscono sia le probabilità a priori delle diverse possibili "cause" di E (ma non si sa per effetto di quale di loro E si è verificato), sia le probabilità condizionate di E data ciascuna delle cause, è possibile calcolare la probabilità che E si sia verificato per effetto di una particolare causa.

Probabilità e densità di probabilità Se Ω è un insieme discreto, la funzione P definita su eventi induce una funzione probabilità $\pi_\omega : \Omega \rightarrow [0; 1]$ definita sugli esiti, che è semplicemente la P sugli eventi costituiti da un singolo esito.

Se Ω è un insieme continuo, si definisce per ogni esito ω una *densità di probabilità* $\pi(\omega)$, come limite del rapporto fra la probabilità P e la misura Δ degli eventi che contengono ω

$$\pi(\omega) = \lim_{\|\Delta(A)\| \rightarrow 0} \frac{P(A)}{\|\Delta(A)\|}$$

dove $A \subseteq \Omega$ è un evento che va via via restringendosi, ma contiene sempre ω e $\|\Delta(A)\|$ è la sua misura.

Variabili aleatorie Una *variabile aleatoria* è una funzione che associa ad ogni esito ω un valore numerico. Per esempio, fissata una soluzione \bar{x} , l'impatto $f(\bar{x}, \omega)$ è una variabile aleatoria.

Funzione di distribuzione La probabilità che una variabile aleatoria abbia valore $\leq x$ è detta *funzione di distribuzione* $F(x)$: è una funzione monotona non decrescente di x .

$$F(x) = P(X \leq x)$$

Di conseguenza $F(x) \geq 0$ per ogni x e valgono i limiti

$$\lim_{x \rightarrow +\infty} F(x) = 1 \quad \lim_{x \rightarrow -\infty} F(x) = 0$$

Inoltre

$$P(a < X \leq b) = F(b) - F(a)$$

Se X è una variabile casuale discreta, ossia ammette una collezione numerabile di possibili valori x_1, \dots, x_n, \dots , la distribuzione $F(x)$ è discontinua e presenta crescite brusche in corrispondenza ai valori possibili; l'entità di ogni salto coincide con la probabilità totale degli eventi in cui X ha quel valore.

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

dove $f(x) = P(X = x)$ è detta funzione di densità discreta di X ed è, per ogni valore reale, la probabilità che la variabile X assuma esattamente quel valore.

Nel caso continuo, la funzione di densità $f(x)$ è il limite del rapporto fra la probabilità che X assuma valori in un intervallo contenente x e l'ampiezza di tale intervallo, e quindi indica su quali valori X tende a concentrarsi.

Di conseguenza, $f(x)$ è la derivata di $F(x)$ e, se X è una variabile casuale assolutamente continua

$$F(x) = \int_{-\infty}^x f(u)du$$

Valore atteso Il *valore atteso* $\mathbb{E}[X]$ di una variabile aleatoria X è definito

- per variabili discrete come

$$\mathbb{E}[X] = \sum_{i=1}^{+\infty} x_i f(x_i)$$

- per variabili assolutamente continue come

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$$