UNIVERSITÀ DEGLI STUDI DI PISA

FACOLTÀ DI ECONOMIA

Tesi di Dottorato in Economia Aziendale

Classe SECS P07

OTTIMIZZAZIONE DEL REDDITO E APPRENDIMENTO COMPUTAZIONALE

Candidato: Nicola Ciaramella

Tutor: Prof. Daniela Mancini

Sommario

INTRODUZIONE

L’azienda è un sistema dinamico aperto verso un ambiente dinamico, che attraverso la gestione mira all’equilibrio economico a valere nel tempo e che tende a portare questo equilibrio al più alto livello possibile nel contesto dato mediante la gestione, intesa come serie continua e potenzialmente illimitata di decisioni e azioni. Il fenomeno azienda e il fenomeno reddito che sta al centro della vita aziendale presentano tali complessità da non poter essere racchiusi in formalizzazioni definitive: se così fosse, la gestione aziendale sarebbe un problema algoritmico, e in tutta evidenza non lo è affatto. Cionondimeno, l’espressione di aspetti delle loro dinamiche in modelli formalizzati è certo importante per approfondirne la comprensione teorica, quanto per aiutare concretamente i decisori aziendali nella pratica della gestione.

Il presente lavoro prende le mosse dalla classica equazione dell’equilibrio economico, che contrappone costi e ricavi evidenziando il reddito e legandolo a quantità e prezzi sia dei fattori acquistati che dei beni venduti. Nella sua semplicità formale, interpretata alla luce dei principi di base della dottrina economico-aziendale, questa equazione contiene i primi elementi per lo sviluppo di una teoria (e di una prassi) dell’ottimizzazione del reddito di impresa con strumenti matematici ed algoritmici.

L’ottimizzazione del reddito in particolari momenti della gestione con strumenti matematici rientra a buon diritto nel campo della Ricerca Operativa e in generale della modellistica matematica. In questo lavoro non abbiamo l’ambizione di dare un contributo in questo settore; tenteremo piuttosto di esplorare una metodologia diversa, ma complementare e compatibile, rispetto ai classici modelli di ottimizzazione matematica, e di mettere in luce alcuni aspetti delle possibilità e dei limiti di tale metodologia.

L’intuizione da cui nasce il presente lavoro può essere sintetizzata così: nella prassi gestionale di imprese non dotate di profonde conoscenze scientifiche, i tradizionali modelli matematici risultano spesso impraticabili, per motivi oggettivi e soggettivi; tuttavia possono essere utilmente affiancati da metodi computazionali di più moderata complessità, che avvalendosi delle attuali tecnologie informatiche possono essere implementati con impegni ragionevoli e con risultati apprezzabili. Tali metodi sono per loro natura euristici, nel senso che le loro basi teoriche sono più intuitive che formalmente rigorose e i loro risultati approssimati e soddisfacente, più che precisi e ottimali.

Proporremo quindi un modello di ottimizzazione della fondamentale equazione di equilibrio economico, modello in generale non risolvibile in modo esatto ma utilizzabile con profitto in modo euristico. Il modello sarà ispirato al principio alla base del *Reinforcement Learning*. È questa una branca di una più ampia disciplina, il *Computational Learning*, che concerne i metodi matematici ed algoritmici con cui un attore computazionale (un robot, un software) può apprendere a risolvere problemi. Nel Computational Learning confluiscono molte metodologie non molti anni fa ancora esoteriche, e oggi comunemente applicate nella soluzione di problemi aziendali, per esempio le *Reti Neurali* o gli *Algoritmi Genetici*, così come il complesso di tecniche oggi noto con il nome collettivo di *Data Mining*.

Sono questi degli strumenti preziosi, di crescente diffusione nelle imprese, e non ne sminuiremo certamente l'importanza teorica e pragmatica; vogliamo però ricordare che questi metodi tendono a presupporre l'esistenza di conoscenza pregressa che in qualche modo è formalizzata in modelli e costituisce la base su cui prendere decisioni. Gli eventi che continuamente conseguono dalla gestione aziendale vengono "assorbiti" in tali modelli, che si arricchiscono senza cambiare sostanzialmente la loro natura. Un esempio può essere il database di una compagnia di trasporto aereo che registra sempre nuovi dati sull'andamento delle prenotazioni su web e con questi dati alimenta un algoritmo di fissazione del prezzo, algoritmo definito da precisi criteri scelti a priori dai progettisti per conto dell'azienda. I dati sono sempre nuovi, la loro interpretazione è prefissata. L'algoritmo osserva l'andamento delle prenotazioni, le confronta con lo storico di mesi e anni precedenti, formula previsioni sull'andamento delle richieste per un certo volo nei prossimi giorni e in base a queste previsioni risolve equazioni che ottimizzano i prezzi. Più la tecnica di previsione è efficace, più i prezzi scelti assicureranno un profitto alto. In sintesi: *questi metodi presuppongono una teoria formale del fenomeno aziendale da governare e ottimizzare*.

Il punto critico in questo procedimento è che gran parte della conoscenza è disponibile a priori, sia in forma di dati nel database storico, sia in forma di ipotesi sul fenomeno prenotazioni all'interno dell'algoritmo. Il principio del Reinforcement Learning è invece di apprendere sempre di più dai nuovi dati osservati senza la necessità di disporre a priori di modelli di elaborazione di questi dati. Il nostro agente computazionale, diciamo per concretezza il nostro programma di ottimizzazione, nasce come *tabula rasa*, senza conoscenza su quello che può aspettarsi dai dati né di come interpretarli e utilizzarli. L'apprendimento automatico avviene tramite interazione con l'ambiente (i viaggiatori che prenotano), è un apprendimento empirico, dall'osservazione. Non ci sono regole predefinite. Un programma ispirato a questi principi osserva il suo ambiente, decide di intraprendere una azione in un repertorio, osserva il risultato di ciò che ha fatto e impara a fare di meglio in futuro. Questo avviene passo dopo passo, per tentativi ed errori, con quello che per gli esseri umani è un apprendimento sul campo, un *learning by doing*.

La motivazione del presente lavoro è che molte aziende affrontano i loro problemi di ottimizzazione del reddito proprio in questo modo. Non diciamo certo che in assoluto non dispongano di conoscenza teorica sul fenomeno reddito nel loro particolare contesto, interno dell'impresa ed esterno del loro mercato-ambiente. Vogliamo solo dire, in modo più relativo, che: 1) questa conoscenza teorica è lungi dall'essere così formalizzata da poter essere incorporata in sistemi software; 2) essa è in ogni caso molto parziale e si inserisce in un contesto in cui le decisioni non possono non essere prese in modo molto più euristico e intuitivo. Questo vale specie per le imprese non grandi e non molto avanzate dal punto di vista scientifico-tecnologico, il che è dopotutto vero per la grande maggioranza delle imprese.

Il modello di ottimizzazione del reddito che proponiamo interpreta l'equazione di equilibrio economico in questo spirito, quasi come fosse un problema di Reinforcement Learning. Per essere realistici, non escluderemo affatto che in realtà l'impresa disponga di modelli computazionali di supporto che incorporano una "teoria del reddito e della sua ottimizzazione": al contrario, in alcuni momenti vedremo come questa evenienza può essere sfruttata per rendere più rapido e più efficace l'apprendimento automatico. Però questa sarà considerata una favorevole evenienza, non un requisito.

Nella trattazione del problema di ottimizzazione dedicheremo ampio spazio a un problema noto come *exploration-exploitation trade-off*, che italianizziamo come *dilemma dell'esplorazione*. Il dilemma consiste nel chiedersi, di momento in momento, di situazione in situazione, in quale misura si deve perseguire il vantaggio immediato e monetizzabile con certezza, e in quale misura si deve invece coscientemente adottare decisioni probabilmente non ottimali, ma che permettono di potenziare il patrimonio conoscitivo in vista di maggiori, ma incerti, vantaggi futuri.

Metaforicamente parlando, stiamo chiedendo a un software di quantificare il budget da investire in Ricerca e Sviluppo; oppure di impostare una politica di sviluppo sostenibile del reddito. Fuor di metafora, si tratta di scegliere non una decisione ottimale, ma una sequenza ottimale di decisioni concatenate. Chiaramente, il problema è estremamente complesso, e in effetti è un tema di ricerca aperto da anni nella letteratura scientifica della Computer Science e delle discipline affini. Nel presente lavoro si avanzano alcune proposte non convenzionali su come affrontare il problema.

Lo stile della trattazione sarà a tratti più formalizzato, fino a descrivere dei veri e propri algoritmi, a tratti più intuitivo e rivolto ai principi metodologici. L'intenzione è di fornire una guida per chi volesse applicare questi principi a problemi concreti: il decisore aziendale è più interessato a capire lo spirito, le potenzialità e i limiti dei metodi che presentiamo, mentre il progettista di sistemi necessita talvolta di una più precisa specifica del significato di alcune affermazioni o di alcuni metodi.

Alcune delle idee qui presentate sono state applicate in esperienze reali, nel settore della pubblicità online, dall'autore del presente lavoro. Non però in forma così sistematica e pura da permettere di trarre conclusioni dotate di vero valore scientifico. Tuttavia, le indicazioni emerse hanno un loro valore, seppure empirico, e hanno permesso se non altro di formulare un ambiente di simulazione dotato di un minimo di realismo nel quale sperimentare alcuni dei metodi qui presentati. I risultati riportati danno indicazioni incoraggianti sul valore di questi metodi e sull'interesse di una proposta originale qui avanzata. Tutto ciò necessiterebbe di una sperimentazione molto più ampia e rigorosa, ma l'intero settore dell'ottimizzazione online presenta questo problema: per evidenti motivi, sia le imprese che producono sistemi di ottimizzazione, sia quelle che li utilizzano, sono molto riservate rispetto alle applicazioni. La letteratura scientifica in merito si basa quindi su elaborazioni teoriche supportate da evidenze empiriche aneddotiche o da sperimentazioni su dati simulati che lasciano ampio margine di arbitrarietà al disegno dell'esperimento, minando la possibilità di generalizzarne i risultati.

Occorre considerare che il settore è relativamente giovane, estremamente dinamico e basato proprio sulla competizione scientifica e tecnologica; questo giustifica da un lato l'interesse che per esso si trova nella letteratura scientifica, dall'altro la difficoltà di reperire risultati rigorosi e non solo aneddotici. Si consideri che il Data Mining, disciplina più matura e più diffusa, ancora oggi soffre dello stesso problema: non sono rese pubbliche le evidenze empiriche che permetterebbero di comparare su benchmark affidabili le diverse metodologie proposte.

I temi trattati in questo lavoro si sovrappongono in parte a quelli della disciplina che va sotto il nome di *Revenue Management*, il governo quantitativo del profitto. Per la verità, il Revenue Management ha origini storiche che ne condizionano visibilmente l'impostazione e la focalizzazione su alcuni aspetti economico-aziendali a discapito di altri. Esso è nato per le esigenze dell'industria del trasporto aereo passeggeri, in seconda battuta per l'industria alberghiera.

I metodi del Revenue Management non sono quelli su cui si concentra il presente lavoro, ma le tematiche sono valide come illustrazione del modello di ottimizzazione qui proposto, ed è istruttivo mettere a confronto i due approcci metodologici, convergenti negli obiettivi ma diversi nell'impostazione culturale.

Lo sviluppo della trattazione si articola come segue.

Nel Capitolo 1 "*Equilibrio economico e ottimizzazione*" si introducono i concetti fondamentali che serviranno a interpretare l'equazione di equilibrio economico e il significato della sua ottimizzazione. Il punto di partenza è quindi nella dottrina economico-aziendale classica: il nostro lavoro è infatti una rivisitazione di alcuni principi fondamentali di tale dottrina in una prospettiva orientata a concetti matematici e algoritmici.

Nel Capitolo 2 "*Decisioni, incertezza e informazione*", si introduce il quadro concettuale in cui inserire il nuovo modello di ottimizzazione dell'equilibrio economico. Si evidenzia la natura probabilistica dell'equazione di equilibrio economico e il ruolo che l'incertezza e l'informazione giocano nella sua soluzione.

Nel Capitolo 3 “*Un modello decisionale di equilibrio economico*” il problema di ottimizzazione del reddito mediante la gestione economico è formalizzato come problema di processi decisionali a molti stadi. La formalizzazione proposta predispone l’applicazione di tecniche ispirate al Reinforcement Learning.

Seguono tre capitoli relativi al Revenue Management classico. Si introducono alcuni problemi tipici e alcuni metodi risolutivi, senza alcuna pretesa di completezza, ma con l’intento di mostrare come il modello decisionale proposta sia in grado di inquadrare molti problemi di Revenue Management classici in un nuovo quadro concettuale ispirato all’apprendimento computazionale.

Nel Capitolo 4 “*Modelli e metodi di Dynamic Pricing*” si presentano elementi di Revenue Management basato sui prezzi. Il problema è come manovrare la leva di prezzo per ottimizzare il profitto della vendita di beni a disponibilità limitata. La trattazione è orientata a esempi ed algoritmi.

Nel Capitolo 5 “*Modelli e metodi di Capacity Control*” si passa all’altra branca del Revenue Management classico, l’ottimizzazione del profitto mediante manovre sulle quantità di beni resi disponibili per la vendita. Anche qui si presentano dei problemi di esempio e degli algoritmi risolutivi.

Nel Capitolo 6 “*Economia del Revenue Management*” si discutono gli aspetti micro-economici del Revenue Management, alla luce dei due capitoli precedenti. Qui la trattazione è meno formale e più discorsiva.

Il Capitolo 7 “*Ottimizzazione nel modello decisionale*”, si tratta in modo alquanto formale il problema dell’ottimizzazione del reddito nel quadro concettuale del Capitolo 3. Mentre là si descrivevano i problemi, qui lo scopo è risolverli in modo algoritmico. Si danno anche indicazioni di aiuto alla progettazione concreta di un sistema.

Il Capitolo 8 “*Generalizzazione dell’informazione*” affronta il problema dello sfruttamento dell’informazione disponibile, qualora essa sia risorsa scarsa, come spesso è il caso nelle applicazioni concrete.

Il Capitolo 9 “*Il dilemma dell’esplorazione*” affronta il problema precedentemente descritto del bilanciamento tra vantaggi certi e immediati e vantaggi incerti futuri. Si descrivono in dettaglio diverse strategie, con precise indicazioni per il progetto di un sistema decisionale automatico.

Il Capitolo 10 “*Metodi avanzati di esplorazione*” presenta due proposte metodologiche per il dilemma dell’esplorazione. Una delle proposte è corredata dei risultati della sperimentazione su dati simulati.

1 EQUILIBRIO ECONOMICO E OTTIMIZZAZIONE

1.1 Il concetto di reddito

"Reddito è l'accrescimento che, in un determinato periodo di tempo, il capitale di un'impresa subisce in conseguenza della gestione"[1].

Tale definizione del reddito, patrimonio comune negli studi di Economia Aziendale, evidenzia gli aspetti essenziali del concetto di reddito[2]:

1. il reddito è una *variazione*;
2. il reddito deve poter essere individuato nel *tempo*;
3. il reddito presuppone la presenza di un *capitale*;
4. il reddito è in relazione causa-effetto con la *gestione*.

Definito come variazione, il reddito può essere positivo, negativo o nullo[3]. Nella nostra trattazione, i valori negativi di reddito saranno considerati come normali possibilità.

Il riferimento al tempo implica che il reddito è relativo a un periodo; la natura di questo periodo dipende dagli scopi della specifica analisi che intendiamo condurre. In linea di principio, la gestione deve essere considerata come un tutto unitario, che si estende dalla nascita alla morte dell'azienda, e quindi esiste un solo reddito nella storia dell'azienda stessa[4]. Ciò non toglie che la vita aziendale possa essere scomposta in periodi, e di fatto lo è sempre, per l'imprescindibile esigenza di rilevare e

---

[1] GINO ZAPPA, *Il reddito d'impresa. Scritture doppie, conti e bilanci di aziende commerciali*, Giuffrè, Milano, 1950, pag. 277

[2] ANTONELLA PAOLINI (a cura di), *Il reddito: nozione e caratteri*, in [MARCHI 2006] pag. 109. In questo capitolo l'esposizione si ispirerà in diversi punti alla trattazione reperibile in questa fonte.

[3] Alcuni autori preferiscono riservare il termine *reddito* alle variazioni positive, utilizzando il termine *risultato economico* nella nostra accezione più generale.

[4] Questo principio è ripetutamente e fermamente enunciato nella dottrina economico-aziendale. Non, ovviamente, per negare l'utilità delle suddivisioni in periodi della gestione, ma per ricordare sempre il loro carattere astratto e strumentale.

comunicare il reddito e altre misure della gestione durante la vita dell'azienda (il bilancio di esercizio è il primo esempio, non l'unico). La scomposizione in periodi può essere puramente convenzionale e decisa dall'analista, per esempio scomponendo la gestione in mesi o trimestri. Può essere convenzionale ma dettata da vincoli esterni, tipicamente la divisione in esercizi richiesta dalla normativa civilistica e fiscale. Oppure può essere legata ai tempi del ciclo economico di produzione[5]. Ciò non comporta violazioni del principio dell'unità temporale della gestione, che non confligge con le tradizionali suddivisioni temporali concretamente usate nelle imprese. La logica profonda del fenomeno reddito resta sostanzialmente la stessa, ed è anzi utile osservarlo su scale temporali diverse e compatibili, quindi con analisi simultanee di breve e di lungo periodo.

Il terzo carattere essenziale del reddito è il riferimento al capitale. Le relazioni fra reddito e capitale sono complesse[6], trattandosi in realtà di due aspetti di un unico fenomeno, la gestione. Un concetto utile per chiarire la questione è la distinzione fra *grandezze flusso* e *grandezze stock*. Gli stock sono grandezze che ha senso misurare a un certo istante, i flussi in un certo periodo. In quest'ottica, gli stock sono visti come il risultato della circolazione e accumulo dei flussi. Il capitale a un certo istante può essere allora pensato come lo stock risultante dalla dinamica del flusso del reddito in un certo periodo, dinamica che risulta dalla gestione. Anche questa metafora "idraulica" è comunque limitata e non può esprimere compiutamente le relazioni fra i due fenomeni.

Nel nostro lavoro useremo anche il concetto di *potenzialità di produrre reddito*. Tale concetto è intuitivamente collegato al concetto di capitale, ma sarebbe eccessivo affermare che ne sia una definizione. Nella dottrina economico-aziendale il capitale è visto come un fondo di valori

---

[5] Il ciclo economico di produzione comprende l'acquisizione dei fattori produttivi, il loro uso nel processo tecnico di produzione, la vendita dei prodotti ottenuti. La durata del ciclo dipende fortemente dalla natura del processo produttivo, e può variare dagli anni alle frazioni di secondo. La mancanza di coincidenza temporale tra i cicli economici e i periodi di osservazione dei risultati fa nascere le problematiche di imputazione dei risultati ai periodi della gestione, e quindi l'introduzione di concetti come ammortamenti, ratei e risconti.

Nella nostra trattazione il concetto di periodo farà di regola riferimento proprio al ciclo economico di produzione, con periodi di durata molto variabile, dettati dall'uso delle tecnologie computazionali, quindi anche dell'ordine dei millisecondi.

[6] Non è corretto pensare che il reddito derivi dal capitale. Anzi, "In senso economico il capitale è prodotto dal reddito; […] non il reddito dal capitale." [ZAPPA] pag. 81.

coordinati, astratti e indeterminati[7]. Si tratta quindi di valori espressi in moneta, astratti in quanto non riferibili a beni concreti, indeterminati in quanto non dotati di un valore misurabile con criteri del tutto oggettivi. Si tratta di attività e passività che, conferite nell'investimento di impresa, costituiscono condizione e strumento della produzione di reddito[8].

Il quarto carattere essenziale del reddito sta nella sua dipendenza dalla gestione, intesa come complesso di operazioni che combinano i fattori e le forze interne ed esterne all'azienda[9]. Nel nostro lavoro la gestione sarà cista essenzialmente come complesso di decisioni, alle quali si suppone seguano le operazioni pianificate. Nella vita concreta delle aziende questa consequenzialità non è scontata, potendo le decisioni restare inattuate, parzialmente o inesattamente attuate. La nostra è peraltro una ipotesi realistica, anche se non completamente certa, in quanto tratteremo di decisioni prese da sistemi informatici i quali, salvo possibili ma rari malfunzionamenti, eseguono esattamente quanto deciso. Inoltre, le decisioni e le operazioni di cui parleremo sono semplici, formalizzate e fortemente strutturate. La nostra semplificazione non inficia quindi la logica di quanto andremo a esporre.

1.2 Il calcolo del reddito totale

Nella pratica aziendale il calcolo del reddito totale è una operazione non usuale, salvo che l'azienda abbia vita brevissima. Noi però andremo ad esaminare fenomeni molto particolari, che possono essere artificialmente separati del resto della gestione aziendale senza inficiarne la natura, in quanto possono essere visti essi stessi come micro-aziende di brevissima vita. Una campagna pubblicitaria su Internet e la vendita di biglietti per un singolo volo possono essere pensati come "progetti"

---

[7] PAOLINI, *Il reddito: nozione e caratteri*, cit., pag. 111-112.

[8] *Ibidem*, pg. 111. Beninteso, condizione e strumento che hanno significato solo in quanto oggetto della gestione, non in sé.

[9] Siamo qui nella visione del fenomeno aziendale propria del pensiero economico-aziendale, che lasciamo sintetizzare alla classica definizione di Giannessi: "[L'azienda può essere intesa come] una unità elementare dell'ordine economico-generale, dotata di vita propria e riflessa, costituita da un sistema di operazioni, promanante dalla combinazione di particolari fattori e dalla composizione di forze interne ed esterne, nel quale i fenomeni della produzione, della distribuzione e del consumo vengono predisposti per il conseguimento di un determinato equilibrio economico, a valere nel tempo, suscettibile di offrire una remunerazione adeguata ai fattori utilizzati e un compenso, proporzionale ai risultati raggiunti, al soggetto economico per conto del quale l'attività si svolge" [GIANNESSI, 1960, p. 46].

autonomi, dotati di un proprio capitale e di un proprio conto economico, con una vita misurabile in giorni. In realtà, a ben vedere, i legami con gli altri aspetti della gestione esistono e nella pratica dell'ottimizzazione della redditività possono talvolta influenzare o anche modificare radicalmente i modelli che andremo a studiare. Ciò non toglie che sia corretto delimitare artificialmente il campo di studio e di applicazione dei modelli, salvo poi confrontarsi caso per caso con le concrete esigenze gestionali che nel modello non possono trovare espressione.

Il calcolo del reddito totale con il *metodo indiretto* consiste nella variazione del capitale fra il momento della costituzione e quello della liquidazione:

Reddito totale = Capitale finale – Capitale iniziale[10]

Con il *metodo diretto*, il reddito viene invece misurato più analiticamente come differenza fra ricavi e costi dell'intera vita aziendale:

Reddito totale = Ricavi totali – Costi totali

Queste formulazioni vengono a coincidere nel risultato finale, purché si usino metodologie coerenti per la valutazione del capitale da un lato e dei costi e ricavi dall'altro[11].

1.3 La competenza economica e il problema dell'attribuzione

Il calcolo del reddito di periodo introduce il problema della valutazione del capitale a un momento intermedio della vita dell'azienda e il problema di attribuzione di costi e ricavi ai diversi periodi. Quest'ultimo è il problema della *competenza economica*. Una voce di costo deve essere attribuita al

---

[10] Questa definizione presuppone la possibilità di misurare omogeneamente i due valori di capitale. In concreto ciò è possibile, perché nei due momenti estremi della vita di azienda il capitale viene ad essere costituito da soli valori numerari. Occorre anche depurare il calcolo da variazioni patrimoniali non dipendenti dalla gestione, siano esse aumenti o diminuzioni di capitale, prelievi di utili o reintegri di perdite. In presenza di inflazione occorre poi calcolare i due valori numerari su base comune, quindi svalutandone uno in base all'andamento dell'inflazione stessa.

[11] Non è questa la sede per affrontare tali problemi, centrali nell'economia aziendale e con vastissime implicazioni che vanno molto oltre le esigenze dei nostri modelli semplificati. Nel testo diamo solo brevi ed ellittici cenni. Due riferimenti che possono chiarire tali concetti a un livello introduttivo, più che sufficiente per il presente lavoro, sono [MARCHI, 2006] e [BIANCHI MARTINI].

periodo in cui si è avuto l'effettivo utilizzo dei relativi fattori nel ciclo produttivo, periodo che non necessariamente coincide con quello in cui il costo si manifesta in forma di entrata o di debito. Per di più, il costo può essere di competenza di più di un periodo. Analoghe considerazioni valgono per i ricavi. Pertanto, l'attribuzione del reddito ai periodi richiede l'attribuzione dei costi e dei ricavi, che a sua volta richiede una analisi delle effettive vicende del ciclo economico nello specifico contesto della gestione di quella particolare azienda in quel particolare momento della sua vita.

Nella vita aziendale "normale" questi problemi vengono affrontati con gli strumenti concettuali dei costi sospesi e ripresi, del consumo dei fattori a lento rigiro, e quindi con i ratei e i risconti, gli ammortamenti e i fondi. Nei nostri modelli questi concetti non appaiono esplicitamente, ma in realtà sono presenti, sia pure sotto forme non convenzionali.

A ben vedere, il problema della competenza economica come attribuzione temporale di quote di reddito a periodi è parte un problema più ampio: l'attribuzione di valore a particolari aspetti del processo gestionale. Dire che un terzo di una voce di ricavo che ha avuto manifestazione lo scorso anno è di competenza dell'anno corrente (e quindi computare un ricavo ripreso all'inizio del nuovo periodo) significa affermare che il "merito" del ricavo è stato per due terzi delle operazioni dello scorso anno e sarà per un terzo di quelle dell'anno corrente. Ciò comporta una articolazione della gestione che sembra contraddire il principio della sua unitarietà. In realtà siamo certo in presenza di un problema, ma non di una contraddizione.

L'azienda è un sistema e quindi deve essere pensata nella sua interezza e nelle correlazioni interne fra i suoi elementi: questo è un principio basilare dell'economia aziendale[12]. La sua gestione è un processo unitario, questo è un altro principio basilare, come abbiamo visto. Ma la stessa economia aziendale provvede il repertorio concettuale per analizzare parti di quel sistema ed elementi di quel processo. Il continuo richiamo all'unità sistematica non nega la possibilità dell'analisi e dell'astrazione, piuttosto è uno stimolo ad assumere una visione sistemica e un memento della natura in parte arbitraria dell'analisi. La teoria della valutazione, la metodologia della chiusura del bilancio, quella della revisione aziendale contemplano tutte la necessità (per imprescindibili esigenze operative) di procedere a suddivisioni più o meno convenzionali, ma pur sempre razionali e motivate.

[12] [BERTINI] è la classica introduzione alla visione sistemica della dottrina economico-aziendale.

Tornando al problema dell’attribuzione del reddito alle parti della gestione cui si può razionalmente attribuire la sua generazione nel ciclo economico, vedremo che questo è un tema ricorrente di modelli di apprendimento computazionale, e che l’attribuzione temporale ne è solo un aspetto.

Per non lasciare del tutto nell’astrazione queste considerazioni, prendiamo un esempio tratto dalla pratica della pubblicità su web e del commercio elettronico. Un’azienda X commissiona campagne pubblicitarie a vari editori elettronici, in pratica a vari siti web. Ciascun editore ospita sul suo sito lo stesso *banner* (l’annuncio pubblicitario); se un visitatore del sito fa un click sul banner, viene portato sul sito dell’inserzionista X; qui vede un’offerta di prodotti e ha la possibilità di acquistarne uno. La formula commerciale prescelta è la cosiddetta *pay-per-sale*: se un visitatore di un sito che ospita la campagna pubblicitaria fa click sul banner e poi compra un prodotto, allora l’inserzionista compensa l’editore con una provvigione, altrimenti non paga niente (e l’editore ha un danno perché ha occupato parte del sito con il banner senza trarne alcun utile, mentre *forse* mostrando il banner dell’inserzionista Y avrebbe potuto averne uno).

L’enunciato sembra a prima vista ragionevole, ispirato a criteri equi e privo di particolari problemi gestionali e contrattuali, a meno di malfunzionamenti informatici. In realtà, ad una indagine più approfondita mostra serie problematiche di attribuzione del reddito alla gestione. Prendiamo il caso di un visitatore che va sul sito A, vede il banner di X (perché il sistema software dell’editore di A ha deciso di mostrargli quello e non il banner di Y), fa click, va sul sito di X e *non* compra alcunché. Il giorni dopo lo stesso visitatore torna spontaneamente sul sito di X e compra un prodotto[13]. Il sistema informatico di X sa che quello è lo stesso visitatore che il giorno prima aveva fatto click sul banner nel sito A[14]. Ci chiediamo se X è tenuto a corrispondere la provvigione ad A. Come è facile immaginare, gli editori dei siti trovano naturale la risposta affermativa: il visitatore è tornato sul sito di X perché “catturato” dalla prima visita, alla quale è da attribuire il merito, e quindi si sono verificate le condizioni contrattuali per corrispondere la provvigione. Non inaspettata anche la riserva sollevata dagli inserzionisti: moltissimi visitatori entrano sul sito spontaneamente, anche senza avere mai fatto click sul banner, quindi *forse* quel visitatore sarebbe entrato lo stesso, perciò

---

[13] “Spontaneamente” nel senso che non viene istradato da un banner su un sito, come il giorno prima, ma scrive esplicitamente l’indirizzo del sito di X, dichiarando la volontà di visitarlo.

[14] Il sistema informatico può essere o non essere in grado di riconoscere che si tratta dello stesso visitatore. Dipende dall’uso o meno della tecnica dei *cookie*, sula quale non ci dilunghiamo. Per il nostro esempio supponiamo che ciò accada.

forse il merito della vendita è della seconda visita, quella spontanea, e non è ovvio che la provvigione debba essere corrisposta. Se il lettore trova speciosa questa riserva ed è portato intuitivamente a parteggiare per l'editore, si chieda se fa differenza il caso in cui la seconda visita segue la prima non di un giorno, ma di una settimana o di un mese. Al crescere dell'intervallo temporale i dubbi aumentano. C'è di peggio: il visitatore fa click sul banner nel sito di A ma poi non compra sul sito di X; dopo una settimana fa click sul banner nel sito di B e poi compra sul sito di X. Il merito è di A, di B o di entrambi? E se il visitatore non compra neanche alla seconda visita, ma dopo un'altra settimana esegue una terza visita spontanea e finalmente compra?

Questi problemi di attribuzione vengono risolti con criteri convenzionali prefissati: vale la prima visita *post-click*, vale l'ultima, in caso di vista spontanea non si paga alcuna provvigione, si paga metà provvigione, si considerano solo le visite post-click della settimana precedente l'acquisto, e simili. Tutte queste soluzioni sono sensate ma non convincenti fino in fondo, e l'intero modello commerciale pay-per-sale stenta a decollare a causa di queste incertezze foriere di contenziosi.

In questo esempio, del tutto realistico ed attuale, il reddito generato dall'acquisto per l'inserzionista X e il reddito generato dalla provvigione per gli editori A e B possono essere attribuiti a diverse operazioni della gestione: da un lato quelle del ciclo economico in cui X incarica A di mostrare *a volte ma non sempre* il proprio banner e A decide *quando e a chi mostrarlo*[15]; dall'altro le operazioni del ciclo economico proprio di X, che pubblica sul proprio sito i contenuti più interessanti per aumentare il numero di visite spontanee. C'è un problema di competenza economica, non tanto dei periodi quanto delle operazioni (risolto questo, si può risolvere più facilmente quello). Il problema non può essere risolto perfettamente, perché il sistema delle operazioni è troppo complesso e imprevedibile; ma non può essere aggirato, pena il blocco della gestione di entrambe le aziende. Occorre perciò trovare una soluzione astratta e imperfetta, ma operativamente ragionevole (oppure rinunciare a questa formula commerciale).

1.4 Il calcolo del reddito di periodo

Il reddito di periodo può in prima approssimazione essere pensato come semplice differenza

---

[15] Sempre con la tecnica dei *cookie*, gli editori possono conoscere l'identità e le preferenze dei visitatori dei loro siti. Non sempre accade, sia per motivi tecnici che etici e normativi.

Reddito di periodo = Ricavi di periodo – Costi di periodo

Questa formulazione è solo un punto di partenza, che per assume valore operativo richiede una metodologia di correlazione fra costi e ricavi e una soluzione al problema della competenza di periodo Il modello del reddito di periodo può essere riformulato in modo più significativo così:

Reddito di periodo = Valore prodotto nel periodo + Variazione valori in rimanenza nel periodo

Variazione valori in rimanenza = Valori in rimanenza a fine periodo – Valori in rimanenza a inizio periodo

Qui c'è da determinare che cosa si intende con *valore*. I prodotti rimasti invenduti a fine periodo potranno essere valutati con criteri di costo (come se ci si attendesse profitto nullo) o di mercato (con una qualche stima del profitto atteso). Si aprono qui le problematiche di costi ripresi e sospesi, di ricavi ripresi e sospesi, di valutazione delle rimanenze. Tale problema nella sua formulazione generale non rientra nei temi di questo lavoro, ma necessitiamo comunque di un modello per rappresentare questo fenomeno.

Enunciamolo dapprima in questo modo:

Reddito di periodo = Flusso netto di periodo + Variazione di stock di periodo

Flusso netto di periodo = Flusso di ricavi di periodo – Flusso di costi di periodo

Variazione di stock di periodo = Stock a fine periodo – Stock a inizio periodo

intendendo che sia il flusso sia la variazione di stock possono essere positivi o negativi.

Riprendiamo il nostro esempio di applicazione alla pubblicità online.

Prendiamo il caso di un editore che suddivide il tempo in ore. In una certa ora il sito ha generato un milione di *impression*, cioè di visualizzazioni di annunci ai visitatori. Queste impression rappresentano un insieme di costi fissi e variabili, per esempio gli ammortamenti degli investimenti in infrastrutture e le quote parte del costo del personale che alimenta e gestisce il sito. Ammettiamo per semplicità che questi costi siano fissi e non eliminabili, e che quindi possiamo per il momento ignorarli, semplificando il ragionamento. Allora il flusso netto del periodo è costituito dal sol flusso di ricavi. Immaginiamo che i visitatori abbiano fatto mille click, con un valore medio di un euro (gli inserzionisti pagano all'editore un certo ammontare per ogni click sui loro annunci). Il flusso di periodo vale quindi mille euro. Fin qui tutto è semplice, ma ora arriva il problema davvero interessante: quanto è variato lo stock? E che cosa significa *stock* in questo contesto?

Se per qualche motivo l'editore è autorizzato a generare non più di 2 milioni di impression al giorno, allora possiamo dire che il suo stock *per quel giorno* si è dimezzato. Questo ci sposta dal periodo ora al più lungo periodo giorno, a conferma che considerare sovra- e sotto-periodi è una necessità logica dei nostri modelli. In realtà non si è dimezzato lo stock, ma lo stock delle rimanenti ore del giorno corrente viste come un unico periodo. Ciò complica già il nostro semplice concetto di reddito di periodo. Un altro punto è: il dimezzamento della misura fisica dello stock (un milione di impression invece di due ancora disponibili) induce un dimezzamento del valore economico dello stock, oppure la relazione è più complessa?

Immaginiamo che l'ora appena passata sia quella fra le 22 e le 23. Un milione di impression ha generato un flusso netto di 1.000 euro. Nell'ora successiva, l'ultima della giornata, avremo a disposizione ancora un milione di impression. Le statistiche sulle visite al sito ci dicono che:

1. È molto probabile che il traffico sul sito, cioè il numero di visite, cali nella prossima ora fra le 23 e le 24 e non raggiunga il milione.
2. È molto probabile che la proporzione di click cali e risulti inferiore allo 0,1% dell'ora precedente (mille click su un milione di impression).

È chiaro che il prossimo milione di impression varrà *probabilmente* meno del milione appena visualizzato e che quindi *probabilmente il valore dello stock residuo è inferiore alla metà di quanto fosse un'ora fa*. Perciò la variazione di stock è di per sé problematica da calcolare in quanto valore economico ed è un valore probabilistico, non deterministico come il flusso appena prodotto.

Si può già intuire dall'esempio il concetto di valore di stock a cui vogliamo approdare: un certo *livello di potenzialità di generazione di flussi*. Nel nostro esempio, il milione di visualizzazioni di annunci ancora disponibili rappresenta una potenzialità di generazione di click, e quindi di reddito in senso stretto. Tale definizione diventerà operativa in termini informatici se saremo in grado di trattare questa potenzialità come un complesso di valori numerici, esprimendola in termini logico-matematici. Ma, come vedremo, anche quando questa formalizzazione non è possibile, e nella pratica spesso non lo è completamente, da questa visione potremo comunque ricavare una metodologia di analisi e decisione con un suo valore sia teorico che applciativo.

A questo punto la nostra esigenza è di formulare un modello di reddito che dia una soluzione razionale a questi problemi:

1. Attribuzione di ricavi a costi e di reddito a periodi e sotto-periodi.
2. Definizione formale di valori di flusso e stock in termini monetari e non.
3. Stima probabilistica di valori, guidata da esperienze e previsioni.

1.5 Attualizzazione di valori

Tradizionalmente il valore teorico di un investimento si calcola sommando il valore attuale dei flussi da esso prodotti in un certo arco temporale e aggiungendo il valore residuo al termine del periodo considerato:

$$W = \sum_{t=1}^{n} F_t (1+i)^{-t} + W_n$$

dove

W = valore attualizzato dell'investimento;

W_n = valore residuo dopo l'ultimo periodo considerato;

F_t = flusso netto di cassa nel periodo t;

i = saggio di attualizzazione.

Questo tipo di formulazione trova applicazione anche nel calcolo del capitale economico[16], con motivazioni intuitive legate al valore del denaro nel tempo. Applicata a fenomeni di brevissimo periodo, come nel caso della pubblicità online, non è invece affatto evidente la sua utilità. Svalutare il reddito ogni minuto è un'operazione matematica di scarso interesse pratico, almeno sembra. Nei nostri modelli, però, tale operazione acquisisce un significato particolare, che qui anticipiamo brevemente. Nei valori di stock considereremo anche valori di conoscenza, diciamo capacità di effettuare predizioni efficaci su cosa accadrà nei prossimi minuti in un sito web. Poiché tale capacità evolve in tempi brevissimi, appunto dell'ordine dei minuti, allora acquista senso

[16] [BIANCHI MARTINI] pag. 27.

attualizzare questi valori a cadenze di minuti. Non si tratterà tanto di attualizzare valori monetari quanto valori di risorse che aiutano a generare valori monetari.

Aggiungiamo quindi ai *desiderata* per i nostri modelli anche il trattamento del valore attuale di flussi futuri.

Il diverso perso assegnato ai flussi tiene anche conto, sia pure indirettamente e in modo schematico, dell'importanza dell'equilibrio finanziario, che è condizione necessaria dell'equilibrio aziendale generale.

1.6 L'equazione di equilibrio economico

L'economia aziendale formula le condizioni di equilibrio economico con una equazione lineare:

$$\sum C + r = \sum R$$

dove:

$\sum C$ = somma dei costi;

$\sum R$ = somma dei ricavi;

r = reddito.

Più analiticamente l'equazione può essere formulata come:

$$\sum_{i=1}^{m} f_i \times pf_i + r = \sum_{j=1}^{n} q_j \times pq_j$$

dove:

f_i = quantità del fattore produttivo i impiegata nel periodo[17];

pf_i = prezzo di costo del fattore i nel periodo;

q_j = quantità del prodotto j generata nel periodo;

---

[17] In questa formulazione è implicito il problema dell'attribuzione temporale: per un fattore a lungo rigiro la quantità impiegata deve essere valutata come quota di ammortamento.

pq_j = prezzo di ricavo del prodotto j nel periodo.

L'equazione di equilibrio può essere formulata in modo dinamico estendendola nel tempo su più periodi:

$$\sum_{t=1}^{T} \sum_{i=1}^{m} f_i^{(t)} \times pf_i^{(t)} + r^{(t)} = \sum_{t=1}^{T} \sum_{j=1}^{n} q_j^{(t)} \times pq_j^{(t)}$$

Normalmente si vede il reddito *r* come incognita dell'equazione. La difficoltà non sta ovviamente nel risolvere l'equazione, ma nell'attribuire valori razionali alle altre variabili. Ancora una volta, la determinazione delle quantità e dei prezzi, sia di costo che di ricavo, è un problema assai complesso, per il quale sia la dottrina che la normativa e la pratica professionale hanno elaborato una serie di metodi e criteri, ognuno con i suoi pregi e i suoi ineludibili limiti. Il problema è quindi una certa *indeterminatezza* della misurazione del reddito, indeterminatezza che non significa affatto arbitrarietà.

Niente impedisce di vedere una qualsiasi delle altre variabili come incognita. Una volta fissato un valore del reddito (poniamo, un valore obiettivo ritenuto congruo), ci si può chiedere quale sia il valore di una delle altre, per esempio quale prezzo di costo debba avere il fattore *i* per rendere possibile quel reddito, assumendo tutti gli altri valori come dati. Si apre quindi la possibilità di simulazioni e analisi *what-if*[18], che gli attuali strumenti informatici di produttività personale rendono agevoli ed efficaci.

1.7 Ottimizzazione di prezzi e quantità

L'equazione di equilibrio economico può essere interpretata come problema di ottimizzazione, ponendo il reddito non come incognita da determinare, ma come variabile da massimizzare nel rispetto di certi vincoli sui valori di tutte le altre variabili.

---

[18] Vedi [MARCHI 2003].

Supponiamo di avere un solo fattore produttivo e un solo prodotto. L'equazione di equilibrio economico diventa:

$$f \times pf + r = q \times pq$$

Vogliamo massimizzare il reddito:

$$\max r = q \times pq - f \times pf$$

Assumendo che tutte le quattro variabili quantità e prezzi siano non negative, la soluzione senza vincoli si trova ponendo a zero *f* e *pf* e assegnando valore infinito a q e pq. Evidentemente ciò non ha alcun senso pratico, perché in realtà esistono vincoli sui valori di ogni variabile e anche sulle relazioni tra variabili. Ci saranno dei limiti inferiori per *f* e *pf* e limiti superiori per q e pq; ma più interessante è l'esistenza di vincoli che legano queste variabili fra loro, per esempio

$$q \leq 2 \cdot f$$

In termini matematici, questo è un problema di ottimizzazione vincolata, argomento su cui esiste una sterminata letteratura scientifica. In questo lavoro ci interessa mettere a fuoco un numero ristretto di problemi di questo genere, studiando metodi per affrontare tali problemi con un apparato matematico semplificato, affiancato da metodi di apprendimento computazionale e da schemi concettuali che facilitano le decisioni. Lo scopo non è solo la semplificazione dei metodi risolutivi, ma la capacità di affrontare l'indeterminatezza che di regola rende ardui (e interessanti) i problemi reali di gestione.

Il prezzo del fattore, essendo un prezzo medio, può dipendere dalla quantità di fattore impiegata, per esempio a causa di condizioni commerciali, come gli sconti quantità. Lo stesso vale sul lato dei ricavi : quantità e prezzi possono (non necessariamente devono) essere interdipendenti. Le quantità di fattore e di prodotto sono certamente dipendenti per motivi tecnico-produttivi. Il problema è realistico, e interessante, proprio per l'esistenza di questi vincoli. In generale, costi e prezzi saranno interdipendenti[19].

[19] È il ben noto *Kreislauf* di prezzi e costi, vedi [GIANNESSI].

1.8 Dimensioni di ottimizzazione

L’equazione di equilibrio economico può essere estesa a comprendere altri aspetti rilevanti per l’ottimizzazione, diversi da prezzi e quantità. Ne introduciamo qui due: il *mercato* e il *tempo*.

Con il termine mercato qui intendiamo un insieme di compratori che presenta *omogeneità interna* e *disomogeneità esterna* rispetto agli altri insiemi di compratori. Queste due caratteristiche devono essere *rilevabili* e *sfruttabili*. Rilevabili nel senso che l’azienda può identificare i mercati e distinguerli l’uno dall’altro, sfruttabili nel senso che questa distinzione permette di prendere decisioni diverse per ciascun mercato e trarre da queste un vantaggio economico. In assenza di questi due requisiti, la distinzione fra mercati perde interesse e possiamo allora vedere il mercato come un tutto unico, senza regioni distinguibili al suo interno. Questo concetto astratto di mercato fa evidentemente riferimento alla classica definizione di *segmento di mercato,* un concetto tra i più importanti del marketing. Tale concetto può svolgere un ruolo cruciale per l’ottimizzazione del reddito e in effetti è al centro dell’attenzione in molti settori dell’economia online, aprendo affascinanti opportunità e al tempo stesso inquietanti dubbi.

Anche il tempo ha un ruolo cruciale in molti problemi di ottimizzazione del reddito. Ci sono analogie tra i due concetti di mercato e di tempo, visti come dimensioni di ottimizzazione. Infatti, il tempo ha un ruolo cruciale nei casi in cui diversi periodi di tempo presentano le caratteristiche di omogeneità interna e disomogeneità esterna. Deve essere possibile rilevare caratteristiche che distinguono un periodo dall’altro, e deve essere possibile e utile prendere decisioni diverse da un periodo all’altro, proprio come per i mercati.

Esprimiamo questi concetti in un formalismo matematico per renderli più definiti[20].

Vediamo una decisione come la fissazione dei quattro valori rappresentato nell’equazione di equilibrio economico: prezzo di vendita, quantità di vendita, prezzo di acquisto, quantità di acquisto:

$$D = \langle P_v, Q_v, P_a, Q_a \rangle$$

Sorge subito una questione: il decisore può fissare liberamente i prezzi, ma non le quantità. Le quantità di acquisto sono limitate dalla disponibilità, e soprattutto le quantità di vendita sono

[20] E anche enormemente semplificati, come sempre accade quando si usa l’astrazione matematica.

limitate dalla domanda, cioè dal comportamento dei compratori. La risposta che adottiamo è che le quantità sono da intendere come obiettivi o limiti: in concreto, il decisore non fissa la quantità che venderà, ma la quantità che metterà in vendita, e che potrà vendere in tutto o in parte in dipendenza dagli andamenti della domanda. Pertanto stiamo qui parlando di *capacità*: la quantità massima di prodotto che potremo vendere, non che venderemo. Lo stesso vale per la quantità di acquisto[21].

Le decisioni aziendali non producono effetti, bensì operazioni, che nell'interazione con l'ambiente esterno producono effetti. È questo un principio essenziale per la corretta interpretazione dei fenomeni che si producono nel sistema di azienda[22]. Noi sintetizziamo l'ambiente esterno nel comportamento della domanda, o meglio nei comportamenti delle domande, al plurale perché consideriamo le dimensioni mercato e tempo. Gli effetti per noi rilevanti sono quelli sul reddito, per cui formalizziamo il reddito come differenza fra costi e ricavi che sono funzioni delle decisioni su prezzi e quantità e della domanda:

$$Reddito = Ricavi(Prezzi, Capacità, Domanda) - Costi(Prezzi, Capacità, Domanda)$$

Questo non è il reddito totale, ma il reddito di un mercato, o di un periodo. Generalizziamo così:

$$Reddito = \sum_{Prodotti} \sum_{Mercati} \sum_{Periodi} Ricavi(Prezzi, Capacità, Domanda) - Costi(Prezzi, Capacità, Domanda)$$

Il nostro problema, nella sua accezione più ampia è:

- individuare o creare i prodotti;
- individuare o creare i mercati;
- individuare o creare i periodi;
- per ogni combinazione prodotto-mercato-periodo individuare le decisioni possibili;
- individuare i vincoli di interdipendenza fra decisioni su diversi prodotti-mercati-periodi;

---

[21] Come vedremo, la capacità di acquisto e il prezzo di acquisto svolgono un ruolo più semplice da definire nel processo decisionale, mentre la capacità e il prezzo di vendita sono le leve su cui si concentra l'attenzione. Ciò non significa affatto che acquisti e costi abbiano importanza secondaria per l'equilibrio economico, al contrario. Soltanto, per l'approccio che seguiamo il nostro tema primario saranno i prezzi e le quantità di vendita.

[22] Vedi [BERTINI] e [MIOLO-VITALI].

- scegliere una decisione per ogni prodotto-mercato-periodo.

tutto questo in modo da ottimizzare il reddito totale.

Si sarà notato che abbiamo parlato di “individuare o creare” i prodotti, i mercati e i periodi. Ciò può suonare un po’ artificioso: la creazione di un prodotto è una decisione strategica, e così quella di un mercato. Di regola, prodotti e mercati sono dati del problema, non leve su cui si può intervenire facilmente. In realtà, in alcuni settori è relativamente facile creare in tempo reale nuovi prodotti e anche nuovi mercati, almeno nel senso che a noi interessa, come vedremo.

L’enunciato è di una complessità scoraggiante, in linea di principio. Nella realtà aziendale, questo problema è sempre affrontato e risolto, più o meno bene, più o meno consapevolmente. Il nostro obiettivo non è risolvere il problema nella sua generalità e in modo completo e perfetto. Vogliamo, molto più modestamente e più realisticamente, definire metodi e modelli per risolvere il problema in pratica, di volta in volta circoscrivendo il campo di applicazione, assumendo ipotesi semplificatrici, trovando soluzioni non ottimali ma efficaci[23].

1.9 Interdipendenza delle leve decisionali

Nella nostra formulazione precedente abbiamo detto che prima si individuano o si creano prodotti, mercati e periodi, poi si scelgono prezzo e capacità per quella combinazione. In realtà non c’è sempre un prima e un dopo nel processo logico di ottimizzazione del redito. Il sistema della gestione, con le sue innumerevoli interdipendenze, non si presenta sempre in questa forma lineare, ma implica intrecci e retroazioni estremamente complessi, che per esigenze operative finiamo sempre per semplificare con modelli che mantengono solo parte della complessità originaria del problema (e del resto proprio questa è la natura dei modelli). Per dare concretezza al discorso, ricorriamo a un esempio di fantasia.

Una compagnia ferroviaria dei tempi della colonizzazione del Nord America fissa il prezzo del biglietto dalla costa Est alla Costa Ovest in 100 dollari. Nessuno o quasi compra il biglietto. La compagnia diminuisce progressivamente il prezzo e scopre, dopo alcuni viaggi a vuoto, che al

---

[23] In effetti il temine *ottimizzazione* può indurre in inganno. A volte lo si usa per indicare la ricerca della migliore soluzione possibile in linea di principio, a volte per indicare la ricerca di una soluzione migliore di quelle già disponibili. La nostra accezione è la seconda, ma non potendo certo usare un neologismo come *migliorizzazione*, usiamo il termine diffuso di ottimizzazione.

prezzo di 50 dollari il numero di compratori sale improvvisamente, il treno si riempie per metà e il viaggio diventa remunerativo. La compagnia, manipolando la leva prezzo, ha scoperto qualcosa di importante sul comportamento della domanda. Il treno è però ancora per metà vuoto: la capacità eccede la domanda, ci sono spazi per aumentare il reddito, anche considerando che i costi del viaggio sono in larga misura fissi e quelli variabili influiscono in misura modesta.

La compagnia tenta ancora esperimenti di prezzo e porta il prezzo del biglietto a 45, 40, 35 dollari, ma senza successo. Il numero di compratori aumenta molto poco e non basta a compensare l'effetto negativo del calo di prezzo di vendita. A livello di 30 dollari, però, c'è un nuovo salto nel numero di passeggeri. A quanto pare, ci sono due soglie a 50 e 30 dollari a cui corrisponde un ingresso in massa di nuovi compratori. Fissando il prezzo a 30 dollari la compagnia può riempire il treno, ma ciò ha un risultato più o meno equivalente a riempirlo per metà con il biglietto a 50 dollari (per il gioco dei costi fissi e variabili). Nasce l'dea di fissare due prezzi distinti, a 50 e 30 dollari, destinati ai due segmenti di clientela che si stanno delineando, i Bassi che non vogliono spendere più di 30 dollari e gli Alti disposti a spenderne 50[24]. Però non possiamo semplicemente proporre le due tariffe a tutti e lasciare che ciascuno scelga, perché anche gli Alti sceglieranno al tariffa bassa: sono disponibili a spendere 50 dollari anziché 30, ma non se possono scegliere. I due segmenti di clientela non sono davvero rilevabili e sfruttabili, per il momento non abbiamo ancora due mercati nel senso che ci interessa.

Però possiamo crearli manovrando le leve di prodotto e prezzo. Possiamo sostituire i confortevoli sedili imbottiti con scomode panche di legno in metà dei vagoni. Questo implica un costo, ma siamo disposti a sopportarlo perché ora abbiamo due prodotti: il viaggio in prima classe e il viaggio in seconda classe. Fissiamo i prezzi a 50 dollari per la prima classe e 30 per la seconda. Adesso i Bassi scelgono la seconda classe, sopportando il disagio. Gli Alti, invece, non lo accettano e si rassegnano a pagare la tariffa più alta. Il treno viaggia a pieno carico, con metà passeggeri di prima classe e metà di seconda, e il reddito è notevolmente aumentato.

La compagnia ha sopportato un costo per "rovinare" metà vagoni, e ad ogni viaggio infligge a metà viaggiatori un disagio apparentemente superfluo. Cosa è accaduto che ha fatto crescere il reddito?

---

[24] Potremmo anche denominarli Avari e Prodighi o Ricchi e Poveri, ma questi nomi suggestivi rischiano di portare l'attenzione sul punto sbagliato. Ciò che conta qui è la disponibilità a spendere, non l'analisi dei fattori che determinano questa disponibilità. Che la differenza fra i due gruppi sia oggettiva, di ricchezza, o soggettiva, di prodigalità, è un tema estremamente importante e interessante per l'analista di marketing, ma non rilevante nel nostro contesto.

La risposta è che, creando due prodotti e manovrando i prezzi, abbiamo creato due mercati, che prima esistevano solo in potenza, e poi manovrando le capacità, metà e metà, abbiamo sfruttato al meglio i due mercati.

Tutto questo è reso possibile da una risorsa chiave: l'*informazione*, o la *conoscenza* se si preferisce. Che esistessero dei prezzi-soglia e che questi fossero di 50 e 30 dollari lo abbiamo imparato sperimentando, e abbiamo anche imparato che le capacità più convenienti erano metà e metà e non altre. Abbiamo sopportato costi per la ristrutturazione di metà vagoni e costi opportunità per i viaggi a carico parziale, ma tutto questo non è stato inutile: erano *costi dell'informazione*, che si ripagano con l'ottimizzazione di viaggi futuri. Questa informazione ci è stata rivelata dai viaggiatori stessi, non volontariamente.

Il nostro esempio di fantasia non è poi completamente immaginario. Quando una compagnia aerea pratica due tariffe, di cui una scontata per chi accetta di far passare il week-end fra andata e ritorno, sta creando due prodotti per due segmenti di clientela diversi. Per certi aspetti ancora più simile è il caso del produttore di software che sopporta costi per creare e mantenere una versione depotenziata del proprio prodotto, per creare i mercati *basic* e *premium* e poi manovrare la leva prezzo per ottimizzare il reddito.

In tutti questi casi, stiamo creando dei mercati e quindi delle domande con caratteristiche diverse, per poi sfruttare le loro differenze per ottimizzare il reddito. Le due leve di capacità e prezzo servono a influenzare la domanda. Nella prassi, si tende a privilegiare l'una o l'altra, diversamente da settore a settore. Ad esempio, le compagnie aeree tradizionali puntano molto sulla capacità, limitando i posti disponibili per la classe economica e riservandone alcuni per le classi ad alto prezzo. Le compagnie *low cost* rinunciano invece a limitare le capacità creando classi di posti, e usano piuttosto la leva di prezzo per differenziare gli acquisti precoci e quelli tardivi (il prezzo sale all'avvicinarsi della data di partenza).

1.10 Ottimizzazione del reddito e Revenue Management

Va da sé che tutte le aziende sono e devono essere sempre impegnate a ottimizzare il proprio reddito, e che questo è un argomento di inesauribile complessità, teorica e pratica. In questo lavoro siamo interessati all'ottimizzazione mediante modelli matematici e algoritmici, argomento certo più circoscritto, ma ancora estesissimo. Per meglio agganciare la trattazione teorica alla prassi aziendale, possiamo fare riferimento a un complesso di metodi e principi che va sotto il nome di *Revenue Management*. Il nome di questa disciplina può fare intendere che comprenda la scienza generale dell'ottimizzazione del ricavo, ma in realtà essa si è sviluppata storicamente come branca della Ricerca Operativa concentrata su un numero circoscritto di problematiche, soprattutto tipiche di due settori: il trasporto aereo e l'ospitalità alberghiera. Si tratta quindi di una disciplina molto caratterizzata storicamente, il che provoca spesso una certa confusione terminologica.

Ad esempio, il fatto che si parli di ricavi (*revenue*) e non di reddito o profitto lascia trasparire una peculiarità dei settori industriali che l'hanno originata: elevati costi fissi e modesti costi variabili, quindi discrepanza quantitativamente minima fra l'ottimizzazione dei ricavi e del reddito. Altri esempi si potrebbero fare per evidenziare come il Revenue Management porti ancora segni evidenti del suo concreto sviluppo storico, che apportano certo concretezza alle tematiche, ma che al contempo ne limitano lo sviluppo teorico. Nel nostro studio, incontreremo problematiche evidentemente affini a ciò che storicamente si chiama Revenue Management, ma anche altre che invece ne divergono, pur essendo in ultima analisi convergenti al comune obiettivo dell'ottimizzazione.

Pertanto, pur essendo noi interessati al concetto generale, ed astratto, di ottimizzazione del reddito, esamineremo a fondo alcune problematiche tratte dal più tradizionale e circoscritto settore del Revenue Management. Al prezzo, modesto, di una certa ambiguità terminologica, avremo così agganci concreti verso problemi ben noti nelle aziende di alcuni settori, e di potenziale grande valore anche per altre che invece non ne son pienamente consapevoli.

Per caratterizzare il Revenue Management, diciamo che esso affronta il problema di ottimizzare i profitti giocando su due leve decisionali, i prezzi e le capacità, intese come quantità di beni resi disponibili per la vendita. Per esempio, un albergo farà Revenue Management decidendo prezzi differenziati per tipologia camere, per periodo dell'anno, per numero di prenotazioni (individuali, di gruppo), per preavviso (prezzo maggiore per la prenotazione a poche ore, minore per la prenotazione con anticipo di settimane). E così come sono differenziati i prezzi, lo saranno le capacità: il revenue manager potrà decidere di accettare prenotazioni con grande anticipo solo per una parte delle camere, riservandosi alcune camere per prenotazioni a breve. In generale, si definiranno criteri in base ai quali una camera potrà in ugni particolare caso essere resa disponibile o non disponibile per la vendita, e se sì a quale prezzo. Come si vede, si tratta di decisioni che gli albergatori hanno sempre preso in modo più o meno informale; la differenza è che nel Revenue Management si ricorre a criteri scientifici e modelli matematici e algoritmici.

Una definizione pregnante è data da Phillips[25]: "*[Revenue Management] refers to the strategy and tactics used by a number of industries (…), to manage the al location of their capacity to different fare classes over time in order to maximize revenue*".

Riguardo alla confusione terminologica, diamo una lista di sinonimi o quasi sinonimi di Revenue Management:

- *Yield Management*: ha lo stesso significato di Revenue Management, ma il termine è più usato nell'industria del trasporto aereo passeggeri.
- *Profit Management*: termine più generale e più corretto, perché estende l'analisi oltre i ricavi, ma in pratica più o meno sinonimo di Revenue Management.
- *Profit Optimization*: si sottolinea l'aspetto di ottimizzazione, suggerendo l'uso di tecniche matematiche.
- *Price Management/Optimization*: ci si concentra sul prezzo come leva decisionale, trascurando la capacità.
- *Dynamic Pricing*: ancora concentrazione sul prezzo, stavolta con enfasi sulla variazione nel tempo.
- *Capacity Control/Allocation*: concentrazione sulla capacità come leva decisionale.
- *Demand Management*: si focalizza come manovrando prezzi e capacità si ottenga una influenza sulla dimensione e struttura della domanda; prezzi più alti tendono a ridurne le dimensioni, vincoli di capacità tendono a richiamare certe tipologie di clienti anziché altre.

Nell'uso comune, questi termini sono quasi sinonimi. Nel seguito utilizzeremo i termini Dynamic Pricing e Capacity Control, ricordando però che si tratta di restrizioni pratiche a una problematica teorica più ampia, l'ottimizzazione del reddito.

---

[25] Vedi [PHILLIPS] pag. 120.

2 Decisioni, incertezza e informazione

2.1 Natura stocastica dell'equilibrio economico

L'equazione di equilibrio economico, nella forma in cui è stata presentata, appare come una normale equazione lineare, che può essere risolta se abbiamo disponibili i dati relativi a tutte le variabili meno una incognita, oppure può essere ottimizzata se una variabile è vista come obiettivo e le altre come parametri regolabili. Questa rappresentazione non evidenzia però la natura decisamente probabilistica dei fenomeni aziendali che l'equazione riassume in forma astratta.

Il soggetto della gestione non può semplicemente scegliere i valori delle variabili nell'equazione. L'impresa è un sistema complesso immerso in sistemi molto più grandi e ancora più complessi, primo fra tutti il mercato in cui opera. Le variabili dell'equazione sono in parte endogene, controllabili dall'impresa, in parte esogene, al di fuori del suo controllo. La gestione non regola direttamente le variabili dell'equazione di equilibrio economico: essa genera delle decisioni che si traducono in operazioni, e dall'incontro fra le operazioni e l'ambiente esterno emergono delle conseguenze solo parzialmente prevedibili, che possono essere anche molto diverse da quelle attese dal decisore.

Lo strumento matematico per rappresentare fenomeni di questo genere è la *variabile stocastica*, o *casuale*: una funzione che, dato lo stesso valore in ingresso, genera di volta in volta uno fra molti valori possibili, ognuno con una certa probabilità, che a sua volta può variare nel tempo. Il comportamento di una variabile stocastica può essere modellato e il modello essere tarato per renderlo conforme all'evidenza empirica, per poi analizzarlo e trarne previsioni più o meno destinate ad avverarsi.

Nell'equazione di equilibrio economico compaiono prezzi e quantità, sia di acquisto che di vendita. L'impresa non è libera di scegliere questi valori a suo piacimento. Se fissa un prezzo di vendita, la quantità venduta si manifesterà in modo stocastico, parzialmente ma non totalmente prevedibile. Solo osservando gli andamenti gestionali l'impresa potrà sapere quali siano le quantità, e questa conoscenza potrebbe arrivare troppo tardi per essere davvero utile a fini di ottimizzazione. Se invece fissa una quantità da vendere, dovrà scoprire empiricamente quale sia il prezzo che realmente consente tale quantità.

Le variabili dell'equazione sono quindi legate da un intreccio di interdipendenze di natura stocastica. Per affrontare tale complessità ricorriamo a

- *Modelli*, che danno una rappresentazione logico-matematica del fenomeno isolando alcuni aspetti e trascurandone altri, in dipendenza dagli specifici obiettivi dell'analisi[26].
- *Euristiche*, procedimenti risolutivi che permettono di affrontare in pratica dei problemi la cui soluzione matematica ottimale non è nota o non è praticabile per eccesso di complessità, mancanza di competenze o di strumenti computazionali adeguati (hardware o software). Le euristiche non garantiscono che la soluzione sarà trovata, né che sarà ottima o anche solo vicina all'ottimo[27].
- *Apprendimento*, nel senso di processi computazionali che apprendono dall'esperienza e migliorano la loro capacità di stima e previsione dei fenomeni che devono analizzare o simulare[28].
- *Adattamento*, nel senso di processi computazionali che apprendono a modificare sé stessi e migliorare il proprio comportamento per conseguire i loro obiettivi.

Con questi strumenti il decisore aziendale ha la possibilità di prendere decisioni "migliori", che con maggiore probabilità otterranno buoni risultati di lungo termine nell'ottimizzazione dell'equilibrio economico.

2.2 Modello fondamentale del problema decisionale

Schematizziamo informalmente un problema decisionale con questi elementi:

- Azioni
- Eventi e probabilità di eventi
- Stati e probabilità di stati
- Utilità e stime di utilità
- Tempo

---

[26] Per i modelli in campo aziendale vedi [MARCHI 2009].

[27] Per una importante categoria di euristiche comunemente usate nei sistemi decisionali, le euristiche di ricerca locale, vedi [HOOS].

[28] Per l'apprendimento automatico, vedi [ALPAYDIN] e [MITCHELL].

In un certo istante un decisore si trova in un certo stato del mondo (l'impresa in congiunzione con il suo ambiente) che gli offre un insieme di azioni alternative. Il decisore sceglie una di queste azioni in base a suoi criteri. In conseguenza della decisione, ma in modo stocastico, accadono certi eventi e il mondo entra in un nuovo stato. Uno stato del mondo ha una certa utilità, che misura la sua desiderabilità, il suo valore per il decisore; tale valore può essere monetario, non monetario o anche misto. All'istante successivo il decisore esamina le azioni alternative offerte dal nuovo stato del mondo e prende una nuova decisione. Il processo decisionale prosegue per un certo numero di istanti, che può essere finito o infinito.

Il decisore sa stimare, in modo non esatto, l'utilità di uno stato del mondo. Il decisore sa anche prevedere, sempre in modo non esatta, la probabilità che una sua decisione per una alternativa determini un certo evento o un certo nuovo stato del mondo.

Obiettivo del decisore è massimizzare il valore del percorso che compie tra i vari stati del mondo durante il tempo assegnato al processo. Il valore del percorso è dato dalla somma dei valori dei singoli eventi più il valore dello stato finale (se il tempo non è infinito). I flussi di valore nel futuro più vicino hanno peso maggiore di quelli nel futuro più lontano.

In questo quadro astratto si cerca di rappresentare alcuni aspetti che la disciplina economico-aziendale attribuisce alla gestione e al reddito, che per noi è l'obiettivo del decisore, o almeno l'obiettivo primario:

1. La durata del processo decisionale è illimitata.
2. Il valore da massimizzare è globale all'intero processo.
3. Le decisioni hanno effetti stocastici, non deterministici.
4. Le stime di probabilità sono soggette a incertezza.
5. Le stime di valori sono soggette a incertezza.
6. Il valore si presenta sia come flusso che come stock.
7. Il valore non è solo monetario.
8. Il valore ha peso diverso secondo il tempo in cui si realizza.

In questo modello concettuale il processo decisionale si articola in queste fasi:

1. Identificazione delle alternative, degli eventi e degli stati.

2. Stima delle probabilità con cui le alternative generano eventi e stati.
3. Stima delle utilità di eventi e stati.
4. Individuazione dell'alternativa preferita.

In termini informali, il processo decisionale poggia su quattro pilastri: *scenari*, *utilità*, *probabilità*, *criteri decisionali*.

2.3 Funzione di utilità

Con il concetto di utilità riuniamo gli aspetti soggettivi della desiderabilità di un evento o stato. L'utilità può essere monetaria o non monetaria. Consideriamo due possibili scenari di perfomance aziendale:

1. profitto di 2 milioni di euro, quota di mercato 10%,
2. profitto di 1 milione di euro, quota di mercato 15%.

Ci sono due criteri di valore: il profitto e la quota di mercato. Per scegliere come più desiderabile l'uno o l'altro dobbiamo fare una operazione mentale di conversione fra obiettivi e chiederci se preferiamo un milione di profitto o un 5% di quota di mercato.

Questa scelta può apparire in contraddizione con il principio per cui lo scopo della gestione è il profitto, ma nel nostro quadro concettuale la contraddizione non sussiste. Prescindiamo pure da questioni di teoria *descrittiva* dell'impresa[29], in cui ci si chiede non come le imprese si dovrebbero comportare, ma come concretamente si comportano. Si può discutere se nella pratica i decisori aziendali sono mossi dall'obiettivo del profitto o da altri ad esempio dal prestigio personale, ma non è questo il nostro interesse. Anche restando nell'ambito di una teoria *normativa* dell'impresa, in cui si discute sul comportamento razionale-ideale che "si dovrebbe tenere", avere obiettivi diversi dal profitto non contraddice la visione del profitto come fine ultimo dell'impresa.

Il nostro quadro concettuale rientra nel paradigma della *razionalità limitata*: il decisore che vuole massimizzare il profitto totale[30] non vede l'evoluzione futura della gestione, non può prevedere con

---

[29] Per una introduizone al "vero" comportamento economico degli esseri umani, vedi [PIATTELLI PALMARINI], [LEGRENZI].

[30] Che è interessato all'equilibrio economico a valere nel tempo di Giannessi.

esattezza i flussi di profitto, ha a sua disposizione degli indicatori incerti di quello che potrà accadere e deve prendere le sue decisioni in condizione di incertezza e approssimazione. IL 5% addizionale di quota di mercato è un indicatore incerto di possibili futuri flussi di profitto che il decisore non sa valorizzare esattamente, ma di tuttavia sa fare una qualche fondata stima. Il decisore vede la quota di mercato addizionale del 5% come una potenzialità di profitto addizionale futuro, perché negli esercizi successivi la quota di mercato del 15% conquistata quest'anno potrebbe portare a maggiori profitti rispetto a una del 10%. Deve allora stimare se questi profitti addizionali, futuri e incerti, valgono più o meno di un profitto addizionale di un milione presente e certo. Certamente questa è una operazione soggettiva, ma non solo soggettiva e non per questo irrazionale. Del resto, le stime di costo sono anch'esse in certa misura soggettive, e ancora di più i piani di azienda, ma ciò non impedisce che possano essere più o meno razionali a priori e più o meno di successo a posteriori.

Quanto detto per la quota di mercato può valere per la soddisfazione dei clienti, le competenze del personale e per gli altri *asset* non monetari né facilmente monetizzabili che pure hanno un ruolo importante nella gestione aziendale.

Ciò è perfettamente compatibile con la natura di fine ultimo del profitto. Possiamo dire che il decisore, assegnando valori numerici ai benefici non monetari con suoi criteri soggettivi, sta usando procedimenti euristici di ottimizzazione e che i decisori che sanno farlo meglio lo devono in parte all'apprendimento basato sulla loro esperienza passata.

Il decisore deve quindi confrontare grandezze di natura eterogenea riconducendole tutte a un termine comune di misura e confronto. Possiamo quindi usare il termine *utilità* per indicare la misura comune che rende confrontabili *desiderata* eterogenei. L'utilità non ha un sua unità di misura naturale: per il modo in cui è costruita è un numero puro, misurato in "punti di utilità" che non hanno significato concreto, né è necessario che lo abbiano. Ma i valori più comuni sono espressi in moneta, e quindi è comodo e non dannoso ricorrere a un piccolo abuso linguistico e dire che un cero evento ha utilità di 100 euro.

Abbiamo visto che un decisore ha un suo proprio modo di decidere i "tassi di cambio" fra utilità di genere diverso come profitto e quota di mercato. Deve, cioè, esprimere dei *criteri di preferenza* fra tipi di utilità. Ci sono anche altri problemi di conversione che deve affrontare. Per esempio, deve scegliere se preferisce 1.000.000 di euro di profitto fra un anno o 1.070.000 fra due anni. Si tratta del noto problema del tasso di sconto e del valore attuale netto. Anche questo è un problema di confronto fra utilità eterogenee, che assume la forma di preferenza intertemporale.

I criteri di preferenza del decisore sono poi essi stessi soggetti a variazione nel tempo. A un anno di distanza, probabilmente il decisore assegnerà pesi diversi a profitto e quota di mercato. Il motivo non sta in questioni di gusto, quanto in un diverso stato del mondo, dell'impresa e del decisore stesso. I suoi obiettivi saranno diversi e *l'informazione a sua disposizione sarà diversa grazie all'apprendimento*. A parità di obiettivi, il decisore avrà maturato nuove convinzioni sul valore di potenzialità di profitto da assegnare a un 5% in più di quota di mercato. Per esempio, può essersi convinto che la maggiore quota di mercato non dà all'impresa quelle opportunità supplementari che l'anno precedente si aspettava. Fermo restando il profitto come obiettivo finale, ora il decisore deluso formula previsioni peggiori, stima meno il valore di stock implicito nella quota di mercato e, *a parità di criteri decisionali* prende decisioni diverse da quelle del'anno precedente. Sono cambiati il suo processo predittivo e la sua funzione di utilità.

Nel nostro esempio abbiamo volutamente introdotto un elemento di imprecisione. Abbiamo detto che il decisore deve confrontare un 5% addizionale di quota di mercato e un milione addizionale di profitto. Una formulazione migliore è che il decisore deve scegliere fra i due incrementi a partire da quei due determinati livelli di profitto e quota di mercato. Se le alternative fossero

1. profitto di 100 milioni di euro, quota di mercato 10%,
2. profitto di 99 milioni di euro, quota di mercato 15%.

è intuitivo che l'operazione interiore di scelta del decisore sarebbe diversa. Si vede che la funzione di utilità è sensibile alle variazioni locali, in altri termini che segue una logica marginalista. La differenza fra 1 e 2 milioni è molto più importante di quella fra 99 e 100 milioni, o almeno così è di fatto quasi sempre per i decisori. Potremmo anche dire che a livello di 99 milioni l'utilità, come funzione della moneta, è molto meno elastica che a livello di 1 milione, e che il tasso di cambio con la quota di mercato è molto più basso (nel senso che scambiare un milione per un 5% è molto più desiderabile).

Di regola, anche se non sempre, i valori hanno una utilità marginale decrescente. Questo fa sì che di regola la funzione di utilità sia convessa (a derivata prima decrescente).

Spesso la funzione di utilità è convessa anche rispetto al rischio. Un decisore cui fosse richiesto di puntare 10.000 euro contro 11.000 a testa o croce verosimilmente rifiuterebbe, almeno così accade in pratica nella maggior parte dei casi. Questo significa che 10.000 euro certi sono preferiti a 21.000 euro con probabilità 50%. Si parla in questo caso di *avversione al rischio*. Il decisore può anche essere *neutrale* o *propenso al rischio*.

Anche se ogni decisore può avere la sua personale funzione di utilità e cambiarla nel tempo, possiamo tenerne conto nei nostri modelli. Basta avere l'accortezza di convertire mentalmente gli importi monetari nell'utilità che essi hanno per quel decisore in quel momento. Perciò, a rigore, noi stiamo parlando di *ottimizzazione dell'utilità*. Ma per comodità espositiva continueremo a parlare di ottimizzazione del profitto, come si fa di solito, di nuovo con un abuso terminologico peraltro innocuo.

Ci si può chiedere come si riesce a determinare la funzione di utilità di un decisore in un certo momento. In sostanza, si interroga il decisore, o il decisore stesso si auto-interroga, proponendo una serie di esempi come quello del profitto e della quota di mercato, oppure del gioco a testa o croce. Raccogliendo dati a sufficienza si arriva a determinare il "listino interiore" dei tassi di cambio fra utilità eterogenee, o la serie dei tassi di sconto intertemporali, o la curva di atteggiamento verso il rischio.

2.4 Criterio del valore atteso

Se scommettiamo sul lancio di una moneta non truccata e puntiamo 1 euro contro 1, diciamo che il *valore atteso* del gioco è 0. Il senso intuitivo è che se giochiamo 1.000 volte, ci aspettiamo di puntare 1.000 volte 1 euro, incassando per 500 volte 2 euro e per 500 volte 0 euro. Il calcolo è

$$\frac{500}{1000}(1+1)+\frac{500}{1000}(0)-\ 1000(-1)=\ 0$$

Se puntiamo 1 euro contro 2 allora il valore atteso diventa

$$\frac{500}{1000}(1+2)+\frac{500}{1000}(0)-\ 1000(-1)=\ \frac{500}{1000}=0,50$$

Ci aspettiamo di vincere in media cinquanta centesimi a lancio.

Se scommettiamo sull'estrazione dell'asso di cuori in un mazzo di 40 carte puntando 1 euro contro 29, il valore atteso è

$$\frac{25}{1000}(1+29)+\frac{975}{1000}(0)-\ 1000(-1)=\ -\frac{250}{1000}=-0,25$$

Su 1000 lanci ci aspettiamo di vincere 25 volte, incassando 30 euro a lancio per un totale di 750 euro, a fronte di una puntata di 1.000 euro. Il saldo è -250 euro su 1.000 lanci, con una perdita media di 25 centesimi a lancio.

Il numero 1.000 non ha alcun particolare significato, indicando solo "un numero di lanci abbastanza alto da permettere di far valere la legge dei grandi numeri". Il valore atteso è dunque il valore medio (vincita o perdita) su un grande numero di lanci.

La formula è espressa con la stessa logica in termini di probabilità, anziché di frequenze:

$$Valore\ atteso=\sum_{Scenari} Probabilit\grave{a}\ scenario\times Valore\ scenario$$

Il criterio del valore atteso dice di scegliere l'alternativa che ha il massimo valore atteso. Nelle scommesse precedenti, l'alternativa "non giocare" ha valore atteso

$$VA=1\times 0=0$$

perché c'è un solo scenario che con probabilità 1 da un risultato di parità, né vincita né perdita.

Il criterio ci dice che nel secondo gioco, quello della moneta 1 euro contro 2, l'alternativa migliore è giocare perché *0,50>0*. Nel gioco delle carte l'alternativa migliore è non giocare, perché *-0,25<0*. Nel primo gioco, quello della moneta 1 euro contro 1, le alternative di giocare e non giocare sono equivalenti, e la scelta è indifferente. Non a caso, intuiamo che questo gioco è equo, non favorevole a uno dei due giocatori. Il valore atteso è anche uno strumento per valutare la giusta posta di una scommessa, il giusto prezzo di ingresso per un gioco di business, o il giusto valore di un investimento.

La teoria del valore atteso non pretende di essere una teoria descrittiva del reale comportamento umano, ma piuttosto una teoria normativa di guida a un decisore razionale. Noi la useremo in questa veste, usando la massimizzazione del valore atteso come criterio di ottimizzazione[31].

Fra le azioni disponibili nello stato corrente cercheremo quindi di scegliere quella che *massimizza il valore atteso totale* su tutta la durata del processo decisionale.

Daremo una formalizzazione matematica di questo criterio nella costruzione del nostro modello di equilibrio economico. Qui anticipiamo informalmente il concetto.

Indichiamo con $P(e|a)$ la probabilità che si verifichi l'evento e posto che si è scelta l'azione a, con $P(s|e)$ la probabilità che si passi nello stato s posto che si è verificato l'evento e, con V_E il valore di un evento e con V_S il valore di uno stato. Il valore atteso VA dell'azione a_1 è

$$VA(a) = \sum_{e \in Eventi} P(e|a) \times [V_E(e) + \sum_{s \in Stati} P(s|e) \times V_S(s)]$$

Il calcolo del valore atteso comprende quindi la stima delle probabilità dei possibili scenari evento-stato che possono conseguire alla scelta dell'azione, nonché la stima dei valori degli eventi e degli stati. Con questo abbiamo in apparenza rappresentato soltanto gli scenari che possono verificarsi al primo passo del processo decisionale. Ma in realtà tutti i possibili cammini che il processo decisionale potrà prendere sono già impliciti nella stima della funzione V_S del valore degli stati. Infatti, mentre la funzione V_E rende conto del valore di flusso che consegue immediatamente dall'azione scelta, la funzione V_S rende conto del valore di stock che contiene la potenzialità dei flussi futuri.

Si pensi a una valutazione del capitale economico con il metodo dei flussi di reddito e del valore residuo in cui ci limitiamo a un solo flusso, inglobando nel valore residuo tutti i flussi di reddito futuri. Non è interessante come metodo di valutazione a priori, ma acquista ben altro senso come metodo di *decisione a stadi*. Nel primo periodo si valuta in qualche modo la funzione V_E per gli eventi possibili nello stato presente e la funzione V_S per i possibili stati successivi. Nel secondo periodo si osserva cosa è accaduto nel primo e si ricomincia a valutare il nuovo processo gestionale, con le stesse regole. La valutazione delle potenzialità future del primo periodo può essere molto

[31] Per una discussione sull'applicabilità della teoria del valore atteso, vedi [KREPS].

approssimativa, ma c'è la possibilità di fare stime migliori nel secondo, che ha *migliore conoscenza* delle probabilità e delle utilità. L'intera impostazione è adattiva e si basa sull'apprendimento in itinere nel processo decisionale.

2.5 Criterio del minimo rimpianto

Il *criterio del minimo rimpianto* è equivalente al criterio del massimo valore atteso e viene talvolta preferito per motivi di convenienza di analisi. Ad ogni stadio decisionale il *rimpianto* di una azione è la differenza fra il valore atteso di quella azione e il valore atteso della migliore azione. Quindi il rimpianto totale è definito come la differenza fra il valore totale del processo come è andato e il massimo valore totale possibile, che si sarebbe ottenuto se durante tutto il processo si fosse sempre scelta l'azione migliore.

Ci sono due notazioni da fare sul valore massimo possibile:

1. Non dipende dalla concreta scelta delle azioni, è un valore a priori, la cui utilità è di strumento teorico di misura delle performance decisionali.
2. Dal momento che le azioni migliori non vengono sempre concretamente scelte, per stimare il massimo valore possibile si deve essere in grado di stimare "che cosa sarebbe accaduto se". In alcuni casi ciò è possibile, in altri no.

2.6 Alberi di decisione

Introduciamo gli alberi di decisione in modo informale con un esempio[32].

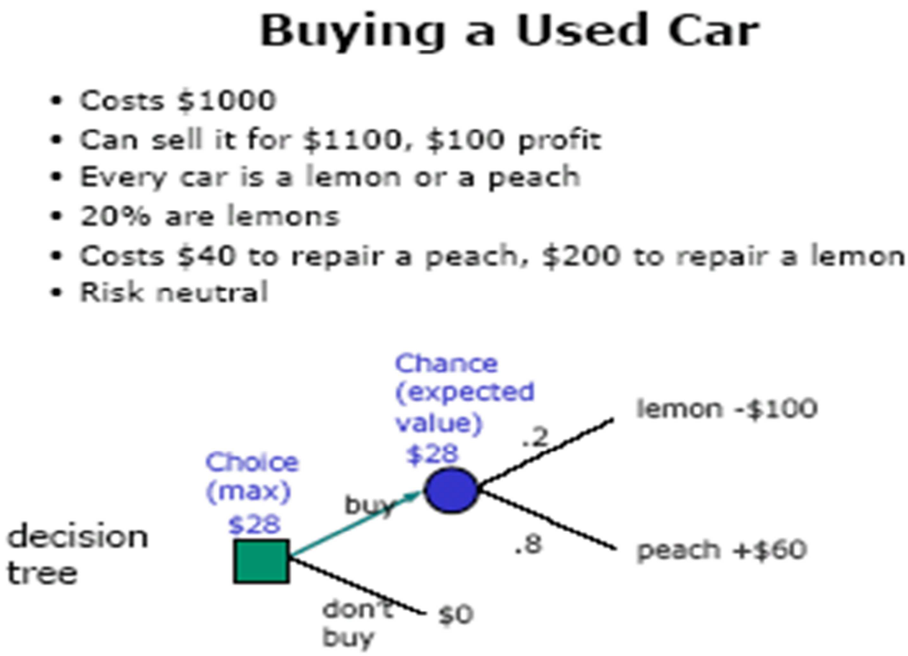
Possiamo comprare un auto usata, con l'intenzione di ripararla e rivenderla. L'auto può essere una *peach* o un *lemon* (termini gergali per auto usata in buono o cattivo stato).

Valutiamo probabilità e utilità per decidere se comprarla o no. Abbiamo un atteggiamento neutrale rispetto al rischio, quindi gli importi monetari sono anche valori di utilità.

Parleremo di questo processo decisionale anche come di un *gioco* per suggerire i concetti di scommessa, posta, speranza di vincita.

---

[32] L'esempio è ripreso da [ANDERSON]. Per una introduzione all'argomento si veda [GHIANI].



Abbiamo la scelta di comprare o no l'auto. Il momento della scelta è rappresentato nell'albero dal *nodo decisione* quadrato. Partono due rami. Quello in basso porta a un *nodo terminale* di valore *0,* il valore atteso di non comprare. Capita spesso che fra le alternative disponibili ce ne sia una neutra di valore nullo. Se scegliamo questa alternativa, il gioco finisce subito. Noi abbiamo fatto una mossa, il mondo non fa alcuna mossa in risposta.

Il ramo verso l'alto corrisponde all'alternativa di comprare, che porta su un *nodo evento* circolare, che rappresenta "la mossa del mondo" in risposta alla nostra. Dal nodo evento partono altri due rami. Il primo corrisponde all'evento "l'auto è un *lemon*", il secondo alla *peach*. Il nodo terminale del *lemon* è etichettato con il valore -100 perché questo è il *pay-off* dello scenario in cui compriamo e troviamo un'auto in cattive condizioni: 1000 dollari di costo d'acquisto, 200 dollari di costo di riparazione, 1100 dollari di ricavo di rivendita. Il nodo terminale dell'auto in buone condizioni ha pay-off +60, dati da costi 1000 + 40 e ricavo 1100.

I rami dal nodo evento ai nodi terminali sono etichettati con le probabilità degli eventi: 20% per il limone, 80% per la pesca. Una volta costruito l'albero di decisione, inseriti i pay-off degli scenari possibili (nodi terminali) e le probabilità degli eventi, si applica il metodo del *folding-back*, il ripiegamento dell'albero.

Nel nodo decisione si inserisce la media dei possibili pay-off pesata con le probabilità:

$$-100 \times 0,20 + 60 \times 0,80 = 28$$

Questo è il valore atteso di gioco con una sola alternativa: comprare. In effetti, il gioco iniziale ha due alternative, ma una volta che ne abbiamo scelta una ci riduciamo a un gioco in cui il mondo fauna mossa e tutto finisce.

Il valore di 28 dollari può anche essere visto come il giusto prezzo d'ingresso per un gioco. Se noi decidiamo di comprare l'auto, e nel momento in cui stiamo per aprire il cofano per ispezionare lo stato dell'auto qualcuno ci propone di sostituirsi a noi nel gioco, assumendosi costi, ricavi e rischi, allora 28 dollari è il giusto prezzo di di *switching* fra noi e il proponente. Un prezzo più alto è favorevole a noi e dobbiamo accettare la proposta, più basso ci è sfavorevole e dobbiamo rifiutare, a questo prezzo la scelta è indifferente. Il motivo è che noi, dopo avere comprato, nel momento in cui stiamo per aprire il cofano *siamo virtualmente in possesso di una vincita di 28 dollari*.

Inseriamo quindi il valore 28 come etichetta del nodo evento. Ora i due rami che escono dal nodo e i relativi nodi terminali non ci servono più per l'analisi: l'informazione che contengono è già sintetizzata nel valore atteso del nodo, che riassume tutto il sottoalbero di cui è radice. Il ripiegamento ha compresso l'albero, perché ha semplificato il problema decisionale.

Adesso abbiamo da considerare il nodo decisione e i due rami che portano sul nodo terminale di pay-off 0 e sul nodo evento di valore atteso 28, il quale ora svolge semplicemente il ruolo di nodo terminale del'albero ripiegato. Qui non abbiamo a che fare con probabilità come per il nodo evento: scegliamo deterministicamente il ramo che ci garantisce il massimo valore atteso, usando il nostro criterio decisionale. Perciò etichettiamo il nodo decisone con il valore 28. Questo valore coincide numericamente con il 28 precedente, ma ha un significato sottilmente diverso: è il valore del gioco con due alternative *se il giocatore può e vuole usare il criterio della massimizzazione del valore atteso*. Se noi possiamo stimare pay-off e probabilità e vogliamo usare il criterio del valore atteso, allora anche nel momento in cui abbiamo *l'opzione* di comprare o no noi siamo virtualmente vincitori di 28 dollari. Quindi questo è il giusto prezzo per entrare nel gioco, equivalente a quello di subentrare dopo che è stata presa la decisione di comprare.

L'albero è etichettato alla radice con il valore 28, che rappresenta il valore atteso dell'intero gioco.

Il nostro esempio descrive un processo decisionale a un solo stadio. Se ci sono più stadi, allora i nodi terminali del primo stadio diventano radici di nuovi alberi che rappresentano i nuovi possibili giochi che si potranno giocare a partire dal secondo stadio, che potranno generare giochi che partono dal terzo stadio e così via. La logica del folding-back è invariata: si parte dei nodi terminali dell'ultimo stadio, si ripiegano gli alberi di ultimo stadio e il nuovo albero è lungo uno stadio in meno. Si prosegue finché l'albero si riduce al nodo decisione iniziale.

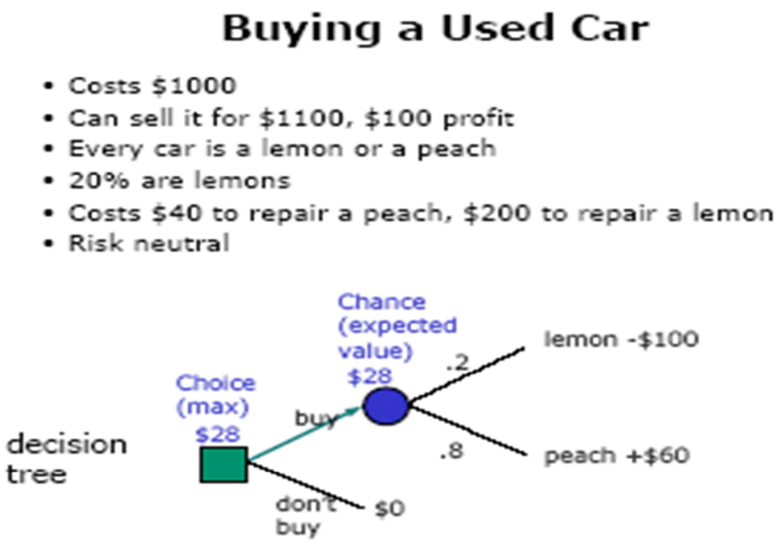
Nel nostro albero il ramo "non comprare" è *dominato* dal ramo "compare", nel senso che non c'è motivo di sceglierlo. Ad ogni stadio decisionale i rami che non hanno il valore atteso massimo sono dominati da quello o quelli che lo hanno (possono essere più di uno). Se nell'albero si eliminano ad ogni stadio i sottoalberi dominati, restano solo i percorsi che descrivono la politica decisionale ottimale.

2.7 Valore atteso dell'informazione

Prendiamo un nuovo e molto semplice esempio di albero di decisone che rappresenta un processo decisionale a un solo stadio.

Possiamo comprare un auto usata, con l'intenzione di ripararla e rivenderla. L'auto può essere una *peach* o un *lemon* (termini gergali per auto usata in buono o cattivo stato).

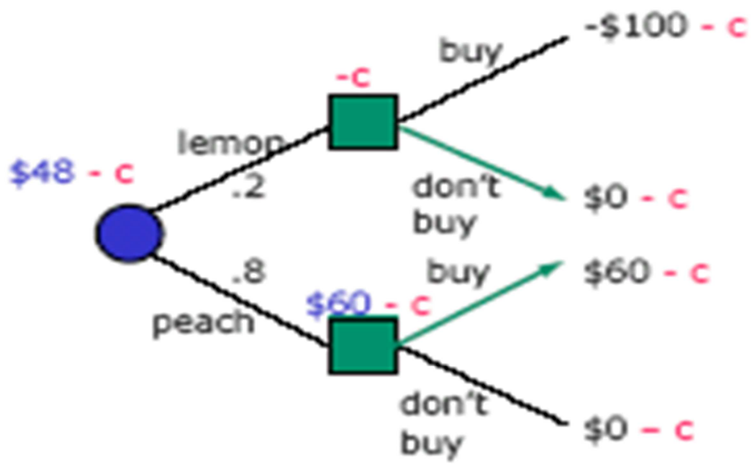
La figura rappresenta l'albero e i dati del problema.



L’albero è etichettato alla radice con il valore 28, che rappresenta il valore atteso del gioco.

Ora immaginiamo che qualcuno, per esempio un dipendente scontento, si offra di dirci in anticipo se l’auto è una pesca o un limone; in cambio ci chiede un compenso. Abbiamo ora un nuovo gioco: possiamo rifiutare l’offerta e giocare il gioco già visto oppure accettarla e giocare un nuovo gioco, considerando anche il costo dell’informazione *c*.

Vediamo l’albero di decisione relativo al gioco con l’informazione.



La radice di questo albero è un nodo evento, perché il gioco comincia con un test che ci rivela se l’auto è un limone o una pesca. Se è un limone, il criterio del valore atteso dice di non comprare, perdendo il costo dell’informazione. Se è una pesca, dobbiamo comprare, guadagnando 60 dollari meno il costo del’informazione. Considerando le probabilità, il valore del gioco è *48-c*.

L’informazione trasforma un gioco di valore 28 in un gioco di valore 48-c, quindi il costo dell’informazione al punto di indifferenza è 20. È razionale pagare meno di 20 e giocare con l’informazione, mentre se il costo è maggiore di 20 allora è razionale giocare senza l’informazione. Diciamo allora che *il valore atteso dell’informazione è 20*.

$$\text{A(Informazione)} = \text{VA con informazione gratuita} - \text{VA senza informazione}$$

Questo semplice modello matematico ci dà un esempio di *investimento finalizzato all’ acquisizione di informazione*: possiamo sopportare un costo allo scopo di acquisire informazione che trasforma i nostri processi decisionali rendendoli più efficaci e creando potenzialità addizionali di profitti.

Parliamo di investimento perché l'acquisto dell'informazione implica una assunzione di rischio, essendo incerto (stocastico) l'effetto finale sul profitto. L'informazione è vista in ottica patrimoniale, come un *asset* intangibile che fa parte del capitale economico in modo implicito, in quanto valore di stock che rende possibili valori differenziali di flusso.

Nell'equazione di equilibrio economico, l'informazione può entrare come fattore produttivo acquistata, ma non appare come quantità venduta, perché in quanto valore di stock essa va ad alimentare non il flusso di reddito presente, ma quelli futuri. L'equazione di equilibrio economico è concepita appunto per rappresentare i valori economici, non anche quelli patrimoniali, come l'informazione. Nasce quindi l'esigenza di collegare l'aspetto economico con quello patrimoniale, se non altro per quel tipo particolare di patrimonio che è l'informazione.

Nell'esempio abbiamo considerato un caso di *informazione perfetta*, in cui l'esito del test è certo: l'informatore ci dice se l'auto è pesca o limone e non abbiamo motivo di dubitare di quanto ci dice. Ora vediamo il caso di *informazione campionaria*, in cui il risultato del test è a sua volta stocastico, incerto.

Un meccanico ci offre di effettuare un test sull'auto, chiedendoci 9 dollari. Il test non è perfetto. Che l'auto passi o fallisca il test non ci dà certezza, solo probabilità.

Per stimare il valore atteso dell'informazione abbiamo ora bisogno di informazioni sul test stesso. Tali informazioni sono rappresentate dalla *matrice di confusione*:

$$\begin{array}{ccc} & Positivo & Negativo \\ Pesca & 0{,}9 & 0{,}1 \\ Limone & 0{,}4 & 0{,}6 \end{array}$$

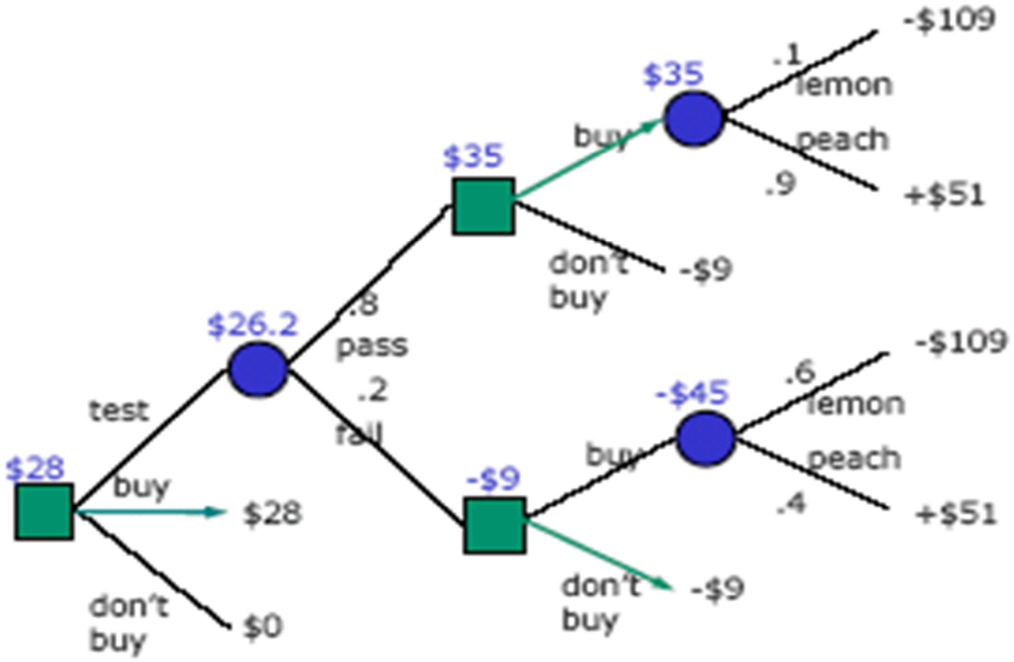
Se l'auto è una pesca, allora il test ha esito positivo con probabilità 0,9 e negativo con probabilità 0,1, mentre se l'auto è un limone le probabilità diventano 0,4 e 0,6.

La probabilità che il test sia positivo è implicita in questi dati e nella nostra conoscenza a priori che le pesche sono l'80% della popolazione di auto e i limoni il 20%[33]:

[33] Usiamo le proprietà delle probabilità condizionate.

e quindi

Ora possiamo costruire l'albero di decisione:



Questo è un esempio di albero decisionale a due stadi. Nel primo stadio abbiamo tre alternative: non comprare l'auto, comprare l'auto senza fare il test, fare il test. Nel secondo stadio, che nasce solo se nel primo abbiamo scelto il test, abbiamo due alternative, comprare o non comprare l'auto.

I pay-off sono valutati tenendo già conto dei 9 dollari di costo per l'informazione. Con il metodo del folding-back calcoliamo il valore atteso degli eventi nel secondo stadio (35 e -45), e poi il valore atteso dei nodi decisione di secondo stadio (35 e -9). Attribuiamo poi il valore atteso all'evento di primo stadio (26,2) e finalmente il valore atteso al nodo decisione di primo stadio (28).

La conclusione è che questo gioco vale 28, come il gioco originario. Il motivo è evidente se teniamo conto dei percorsi dominanti nell'albero (le sequenza dovute a decisioni ottimali per il criterio del massimo valore atteso): l'alternativa ottimale al primo passo è comprare senza fare il test, quindi questo nuovo gioco, se giocato razionalmente, al primo passo si riconduce al vecchio gioco. Se abbassiamo il costo del test sotto i 9 dollari, arriviamo a un punto in cui acquistare o no il test diventa indifferente: è il punto in cui il valore atteso del ramo "fare il test" è a sua volta 28, che si dà il caso valga 7,2. Questo è il *valore atteso dell'informazione campionaria*, il prezzo di indifferenza del test.

Questo esempio ha mostrato che, almeno in linea di principio, ha senso attribuire un valore economico all'informazione, vedendola come un fattore da acquisire che genera un potenziale differenziale di reddito. Tradotto in termini concreti, il problema che stiamo delineando è *valutare il giusto livello di investimento in test che migliorano le nostre capacità di stima delle probabilità di eventi.*

2.8 Informazione e incertezza

Immaginiamo di partecipare a un gioco in cui dobbiamo predire se il colore della pallina che sarà estratta da un'urna contenente palline bianche e nere e incassiamo un euro per ogni previsione corretta. Possiamo scegliere tra due urne quella con cui giocare: la prima contiene 60 palline bianche e 40 nere, la seconda 10 bianche e 40 nere. Con quale preferiamo giocare?

La risposta che diamo è: con la seconda, perché è più squilibrata. Con la prima urna è bene scommettere sul bianco, con la seconda sul nero, ma la scommessa "nero-seconda" è migliore della "bianco-prima", come è intuitivo e come conferma il criterio del valore atteso. La scommessa con la seconda urna è *meno incerta* in senso intuitivo. Oggettivamente, se scegliamo la seconda urna possiamo guadagnare di più (a condizione di giocare razionalmente); soggettivamente se scegliamo la seconda la decisione è più facile e rassicurante, meno ansiogena e soggetta a rimpianti.

Le urne meno incerte possibili sono quelle con tutte palline bianche o con tutte palline nere; le urne più incerte sono quelle con metà palline bianche e metà nere. Non è rilevante il numero di palline, ciò che conta è lo squilibrio interno. Fra un'urna con palline bianche e nere in pari numero, e un con palline rosse, verdi e blu in pari numero, la prima è meno incerta. Il numero delle palline non conta, il numero dei colori sì.

Questa definizione intuitiva dell'incertezza è catturata dalla famosa definizione di Shannon:

$$H(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Se un evento può dare *n* risultati, l'incertezza associata all'evento è la somma dei prodotti di probabilità e rispettivi logaritmi in base 2[34].

Il lancio di una moneta può dare due eventi con probabilità $\frac{1}{2}$ è la sua incertezza è

$$H\left(\frac{1}{2}, \frac{1}{2}\right) = -\left(\frac{1}{2}\log_2 \frac{1}{2} + \frac{1}{2}\log_2 \frac{1}{2}\right) = 1$$

Si dice che l'incertezza di questo evento è di 1 bit. Il senso è che per descrivere il risultato dell'evento serve appunto 1 bit, inteso come numero binario lungo una sola cifra. Stabilita una convenzione, 0 per testa e 1 per croce, o viceversa, ogni lancio può essere descritto con 0 oppure 1.

L'evento "lancio di due monete in ordine" ha 4 possibili esiti e, come si può facilmente verificare

$$H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) = 2$$

Di nuovo, questi esiti si possono rappresentare con un messaggio binario lungo due cifre. Con 3 cifre si rappresentano 8 possibili esiti, con 4 cifre 16 esiti e così via.

Per il lancio di un dado a 6 facce 2 bit sono pochi e 3 sono troppi. Ma in un codice binario codificato in maniera ottimale serviranno a volte 2 bit e a volte 3, e in media serviranno proprio

$$H\left(\frac{1}{6}, \frac{1}{\dots 6}\right) = \log_2 6 \approx 2{,}58\ bit$$

Si può verificare che

$$H(60{,}40) > H(40{,}10)$$

a conferma del nostro esempio intuitivo.

---

[34] Il segno meno serve solo a ottenere un risultato positivo, perché le probabilità sono numeri minori di 1 e quindi hanno logaritmo negativo. Se la probabilità è proprio 1 allora il suo logaritmo è 0. Se al probabilità è 0 allora il suo logaritmo è indefinito, e per convenzione in questa formula lo si considera come se fosse 0. Ne consegue che contribuiscono all'incertezza solo i risultati né certi né impossibili.

Una interpretazione dell’incertezza così definita è la lunghezza media di un messaggio che descrive gli esiti dell’evento in un codice ottimizzato[35]. Intuitivamente, *l’incertezza misura la complessità intrinseca della descrizione dell’evento*, non semplificabile con una migliore scelta di codice.

Poniamo di voler indovinare qual è stata la carta estratta da un mazzo di 40, ponendo una sequenza di domande a risposta sì/no. Possiamo porre la domanda “è il fante di picche?” e sperare di indovinare al primo colpo, ma se va male rimaniamo con 39 carte sui indagare. Oppure possiamo chiedere “è rossa o nera?”, rinunciando a indovinare subito ma ottenendo con certezza la riduzione a 20 delle alternative rimanenti. La strategia ottimale è proprio questa delle domande “dimezzanti”[36], e giocando ripetutamente in media si indovinerà la carta in $\log_2 40$ colpi (un po’ più di 5).

Emerge un aspetto economico dell’incertezza: essa rappresenta il costo di descrivere, o di predire se si preferisce. E se si interpretano i logaritmi della formula appunto come costi, allora la formula stessa appare come un valore atteso, il valore atteso del costo della predizione.

Arriviamo ora al legame tra incertezza e informazione. Si tratta dello stesso concetto visto in modo complementare: *l’informazione riduce l’incertezza*. L’enunciato non è banale, perché essendo l’incertezza misurabile, anche l’informazione lo è: la misura di informazione di un messaggio si misura come riduzione dell’incertezza sul’oggetto del messaggio.

La domanda “è rossa o nera” riduce con probabilità 1 l’incertezza da *H(40)* a *H(20)* e il suo contenuto informativo è di *H(40) – H(20)* bit. La domanda “è il fante di picche?” riduce con probabilità $\frac{1}{40}$ l’incertezza a *H(1)*, cioè a *0*, e con probabilità $\frac{39}{40}$ a *H(39)*. Perciò *il guadagno atteso di informazione* è

$$\frac{1}{40}[H(40) - H(1)] + \frac{39}{40}[H(40) - H(39)]$$

---

[35] I codici binari sono i migliori perché sprecano meno cifre di quanto fanno i codici decimali o in altra base numerica. Questo spiega perché il logaritmo è in base 2. Il riferimento ai codici non è causale: la formula di Shannon nacque per risolvere problemi di codifica e trasmissione di segnali.

[36] Il logaritmo è in base 2 perché le domande della strategia ottimale riducono le alternative rimanenti di un fattore 2.

2.9 Perché l'informazione è desiderabile?

Abbiamo visto nell'esempio dell'auto usata che una quantità addizionale di informazione comporta una guadagno di valore atteso di un processo decisionale, quindi un vantaggio di natura economica. Il motivo è che con più informazione siamo in grado di stimare meglio le probabilità. Esaminiamo meglio questo concetto.

Nell'esempio le probabilità a priori di pesche e limoni erano 0.8 e 0.2. Il test ha probabilità 0.8 di avere esito positivo e 0.2 negativo e nei due casi porta le probabilità a 0.9 e 0.1 oppure a 0.4 e 0.6.

L'incertezza prima del test è

$$H(0.8, 0.2) = 0.72$$

L'incertezza attesa dopo il test è

$$0.8 \times H(0.9, 0.1) + 0.2 \times H(0.4, 0.6) = 0.57$$

Il test ha aumentato lo squilibrio nel caso di esito positivo, il più probabile, a costo di una diminuzione di equilibrio nel caso di esito negativo, meno probabile. Il saldo è un aumento di squilibrio complessivo e quindi una riduzione di incertezza. Il miglioramento di capacità predittiva consiste in questo.

Si provi la seguente interpretazione. Prima del test il futuro del gioco era una entità unica, che si sarebbe rivelata solo in modo spontaneo, a decisione ormai presa. Il test ha diviso il futuro in due futuri separati: il futuro del test positivo e il futuro del test negativo, due mondi virtuali paralleli. Questi due mondi hanno due diverse funzioni di risposta (pesca o limone) alle nostra politica (comprare o non comprare) e noi possiamo sfruttare la loro separazione per attuare due politiche differenziate per i due mondi (comprare se positivo, non comprare se negativo)[37].

Ora descriviamo la segmentazione di mercato in questi termini: suddividere un mercato unico in segmenti che hanno diverse funzioni di risposta (acquisto o non acquisto) alla nostra politica

[37] L'albero mostra che queste sono le decisioni che dominano le loro alternative.

(offerta di un certo prodotto o prezzo anziché un altro) e noi possiamo sfruttare la loro separazione per attuare politiche differenziate per segmento.

In entrambi i casi è desiderabile che i "mondi separati" siano poco differenziati al loro interno e molto l'uno rispetto all'altro. Nel nostro esempio il mondo del test positivo e quello del test negativo sono diversi e al loro interno mostrano un comportamento predominante (più pesche nel mondo positivo, più limoni in quello negativo). La separazione ha quindi un certo valore economico. Che questo sia moderato (7,2 dollari su un gioco che ne vale 28) è conforme all'intuizione che la separazione è solo moderatamente efficace: i due mondi sono diversi ma non moltissimo, il mondo positivo ha sì una forte predominanze di pesche, ma il mondo negativo non ha una forte predominanza di limoni.

Fin qui abbiamo visto una utilità *oggettiva* dell'informazione. Avanziamo ora una ipotesi: l'esistenza di una utilità *soggettiva* e misurabile. È solo una ipotesi speculativa per la quale non abbiamo un supporto di dati empirici. Dal punto di vista dell'esperienza quotidiana questa ipotesi appare alquanto conforme all'esperienza, almeno in prima approssimazione.

Abbiamo detto che la teoria della massimizzazione del valore atteso ha per noi un valore normativo, non descrittivo: ci dice cosa un decisore razionale ideale dovrebbe fare, non cosa i decisori reali fanno davvero. L'esperienza suggerisce che le persone non lo seguono, non in forma pura, o almeno sembra che non lo facciano[38]. Una ipotesi è che in realtà esse seguano i criteri dell'utilità attesa più di quanto sembri, ma inserendo fra le utilità qualcosa di non immediatamente evidente. La nostra ipotesi è che il criterio della massima utilità attesa sia in genere usato con enfasi completamente rivolta all'output delle decisioni, trascurando il loro input. Un decisore umano "sente" il valore atteso come *ricavo* del processo decisionale ma "sente" anche l'incertezza come *costo* di tale processo. È una ipotesi sensata: l'incertezza rende la decisione più faticosa, lenta, ansiogena e perfino più attaccabile con il senno di poi dai critici. In sintesi, è psicologicamente costosa. L'esperienza di vita nelle organizzazioni mostra (parliamo sempre a livello aneddotico e non scientifico) che spesso le scelte facili da prendere hanno il sopravvento su quelle più razionali. Talvolta questo è dovuto a funzioni di utilità perfettamente sensate dal punto di vista del decisore,

---

[38] Vedi [PIATTELLI PALMARINI] e [LEGRENZI].

benché non dichiarabili pubblicamente. Ma spesso sembrano dovute semplicemente a un rifiuto dell'incertezza perché psicologicamente costosa. I decisori gradiscono le decisioni ad alto valore atteso, ma non quelle ad alta incertezza.

Potremmo quindi immaginare una funzione obiettivo che non sia

$$\max_{d \in Decisioni} VA(d)$$

ma piuttosto

$$\max_{d \in Decisioni} \alpha VA(d) - \beta H(d)$$

o forse

$$\max_{d \in Decisioni} VA(d)^{\alpha} \times H(d)^{-\beta}$$

con $\alpha, \beta > 0$.

Abbandonando la speculazione, il presente studio intende fornire strumenti operativi per la progettazione di sistemi informativi decisionali, parzialmente o totalmente automatizzati. In tale contesto, usando un approccio come quello che proporremo nel presente lavoro, è perfettamente possibile sottoporre a verifica empirica ipotesi come quelle che abbiamo appena formulato. Basta modificare la funzione di utilità obiettivo, operazione che dal punto di vista del progetto software è semplice, e osservare cosa accade alle performance oggettive del sistema e alla sua accettazione per motivi soggettivi da parte dei decisori umani che utilizzano i sistemi.

3 UN MODELLO DECISIONALE DI EQUILIBRIO ECONOMICO

Presentiamo dapprima un modello astratto di processo decisionale per l'ottimizzazione del reddito, che poi istanziamo in alcuni modelli applicativi con contenuti più concreti.

Il modello descrive una serie di decisioni distribuite nel tempo. In ogni momento la decisione opera sulle leve di prezzo e quantità, basandosi sulla capacità produttiva e sul patrimonio informativo

Il modello ha carattere stocastico: le decisioni sono prese in base a considerazioni probabilistiche, e le loro conseguenze sono a loro volta prevedibili in modo probabilistico.

Discutiamo le ipotesi su cui il modello si basa e le caratteristiche richieste a un metodo di ottimizzazione basato sul modello. I metodi di ottimizzazione saranno presentati nel capitolo successivo.

3.1 Modello astratto

Abbiamo a disposizione un orizzonte temporale di *T periodi*. In ogni periodo il decisore tiene conto di:

- uno *stato*, che descrive i valori di stock che costituiscono la sua potenzialità di produrre reddito;
- un insieme di *azioni*, azioni che provocano una risposta dell'ambiente esterno.

Il processo decisionale consiste nel prendere in ogni periodo una *decisione* che consiste nello scegliere una azione tra quelle disponibili nel periodo.

Una decisione provoca una risposta dell'ambiente e la combinazione ha due effetti:

- la generazione di un *flusso* di reddito;
- la *transizione* a un altro stato.

Sia il flusso di reddito che la transizione di stato sono funzioni stocastiche dello stato di partenza che della decisione, non sono quindi univocamente determinati, ma si possono assumere di volta in volta valori diversi.

La storia di un processo, cioè il modo in cui effettivamente si realizza, si può descrivere con questa sequenza:

$(S_1, a_1, r_1), (S_2, a_2, r_2), \ldots, (S_{T-1}, a_{T-1}, r_{T-1}), (S_T, a_T, r_T)$

Nel primo periodo il processo è in stato S_1. Il decisore sceglie l'azione a_1, l'ambiente esterno dà una risposta e, in modo stocastico, si genera il flusso r_1 e si passa allo stato successivo S_2, in cui il ciclo si ripete.

Il problema di ottimizzazione si può enunciare così: scegliere le azioni $a_1, \ldots, a_T$ in modo da massimizzare la somma dei flussi.

Le decisioni dei singoli periodi hanno due conseguenze stocastiche, il valore del flusso e lo stato successivo della transizione. Non si tratta quindi solo di generare il massimo flusso o il migliore stato, ma la combinazione ottimale di flusso e stato.

Per tenere conto dell'aspetto finanziario del processo decisionale, introduciamo il *tasso di sconto* $0 \leq \gamma \leq 1$ che svaluta i flussi futuri. Il nostro problema si pone come

$$\max_{a_1,\ldots,a_t} \sum_{t=1}^{T} \gamma^{t-1} r_t$$

Per descrivere il problema di ottimizzazione abbiamo bisogno delle funzioni stocastiche delle transizioni di stato e dei flussi:

$$\sigma(s,a,s') = \Pr\{s_{t+1} = s' | s_t = s, a_t = a\}$$

$$\rho(s,a,s') = \mathrm{E}\{r_t | s_t = s, a_t = a, s_{t+1} = s'\}$$

Se lo stato di attuale è s e l'azione selezionata è a, la funzione σ restituisce la probabilità che si passi allo stato s' e la funzione ρ restituisce il valore atteso[39] del flusso in dipendenza del nuovo stato.

[39] Il valore atteso è la media dei possibili valori pesata con le rispettive probabilità. Se passando dallo stato s allo stato s' eseguendo l'azione a ottengo con probabilità 2/3 un flusso di valore 6 e con probabilità 1/3 un flusso di valore 15, allora $\rho(s,a,s') = \frac{2}{3}6 + \frac{1}{3}15 = 9$.

Dalle definizioni si vede che lo stato di partenza e l’azione selezionata determinano congiuntamente il flusso e il nuovo stato, che non sono indipendenti.

Le funzioni ρ e σ non sono note, ma possono essere progressivamente apprese durante la storia del processo osservando gli effetti delle decisioni.

Resta da definire come in ogni stato sarà scelta l’azione. Introduciamo quindi il concetto di *politica* come funzione stocastica dello stato per la scelta di una azione: $\pi(s, a)$ è la probabilità di scegliere l’azione a nello stato s secondo la politica π.

Riassumendo, la specifica completa di un modello deve dare un significato concreto a questi elementi:

- S, gli stati;
- A, le azioni;
- σ, la funzione di transizione di stato;
- ρ, la funzione di generazione di flusso;
- γ, il tasso di sconto;
- π, la politica di scelta.

Le definizioni delle funzioni di stato e di flusso dicono che le conseguenze delle decisioni dipendono in modo stocastico solo dalle decisioni stesse e dagli stati, ma non dalla storia del processo fino al momento. Questa è la *proprietà di Markov*, che circoscrive una classe importantissima di problemi decisionali, classe in cui rientrano quelli che ci interessano in questo lavoro. Un processo decisionale markoviano ha un unico parametro, lo stato definito in qualsivoglia modo, che fornisce tutte le informazioni necessarie per le decisioni, rendendo superflua la storia passata a fini decisionali. Ciò significa che una politica di scelta ha come argomento solo lo stato e niente altro. Un esempio di processo markoviano è il gioco degli scacchi. Il giocatore che deve muovere trae tutte le informazioni le informazioni necessarie dalla posizione dei pezzi, senza bisogno di conoscere la sequenza di mosse che ha portato fin qui. Pertanto gli scacchi sono un gioco markoviano, ammesso che lo stato sia definito appunto come la posizione di pezzi. Potremmo anche definire lo stato come la sequenza di tutte le mosse, e in questo caso avremmo ancora un processo markoviano, ma con un numero enormemente superiore di stati[40]. Per essere rigorosi, la posizione di pezzi non contiene proprio tutta l'informazione necessaria. Poiché si può giocare la mossa dell'arrocco una sola volta nella partita e a condizione che il Re non sia stato precedentemente mosso, il giocatore ha bisogno di sapere se le condizioni dell'arrocco sono soddisfatte nel momento in cui esamina la situazione. Questa informazione storica può essere inserita anch'essa nello stato, che diventa quindi la combinazione di posizione dei pezzi e diritto o non diritto di arrocco.

L'esempio degli scacchi chiarisce che la proprietà di Markov è del modello, non del fenomeno modellato. Se lo stato contiene le informazioni utili per la decisone, allora è markoviano. Quali siano le informazioni da inserire nello stato lo decide il progettista del modello. Non si deve pensare che uno stato debba contenere solo informazioni visibili al presente nel fenomeno reale. Si può inserire nello stato una traccia del passato, come il diritto di arrocco, che da quel momento diventa una variabile presente a tutti gli effetti. Le variabili di stato sono la memoria della storia del processo. Poiché l'utilità del modello è una proprietà graduale, non del tipo sì o no, è perfettamente possibile omettere dallo stato molti aspetti della storia e ottenere ancora un modello soddisfacente del fenomeno reale, guadagnando in semplicità.

I modelli markoviani hanno proprietà matematiche e algoritmiche utilissime, che li rendono strumenti pratici preziosi. Per questo motivo i progettisti di modelli tendono a forzare la proprietà markoviana appena possibile, accettando una certa mancanza di accuratezza del modello in cambio di un trattamento molto più agevole.

Per concretezza diamo alcuni esempi di varia complessità di modelli derivati da questo specificando la natura degli elementi astratti stato, azione, ambiente, risposta, transizione e flusso.

3.2 Contesto di esempio

Abbiamo un portafoglio di beni alternativi X_1, X_2, …, X_n che intendiamo vendere online. Abbiamo affittato per 24 ore uno spazio espositivo su un sito di e-commerce. Il nostro sistema informatico comunica con il server del sito e ogni ora invia queste informazioni:

---

[40] Per la verità anche le posizioni dei pezzi sono un numero enorme, ma quello delle sequenze di mosse è enormemente superiore a questo numero enorme.

- *Prodotto*. Quale prodotto promuovere nello spazio espositivo.
- *Prezzo*. Quale prezzo offrire al visitatore del sito, il nostro potenziale compratore.
- *Capacità*. Quante unità di prodotto vendere al massimo in un’ora.

Lo spazio contiene soltanto una descrizione di prodotto, non più di una. Il prezzo è lo stesso per tutti i visitatori. Se la capacità è esaurita prima di un’ora, per il resto dell’ora rifiutiamo la vendita.

Al termine dell’ora il server del sito ci comunica quante unità di ciascun prodotto sono state richieste dai visitatori; siamo quindi in grado di calcolare il ricavo. Non ci sono ritardi di registrazione per vendite effettuate subito prima di mezzanotte. Inviamo di nuovo le nostre scelte di prodotto, prezzo e capacità e il ciclo orario ricomincia. Abbiamo quindi un problema decisionale a 24 stadi.

Il nostro obiettivo è massimizzare il ricavo. In realtà vogliamo massimizzare il margine di contribuzione, ma per semplificare l’esempio ignoriamo i costi e quindi margine e ricavo coincidono. Non consideriamo il tasso di sconto, che poniamo uguale a 1.

La domanda è *funzione stocastica* della nostra decisione su prodotto, prezzo e capacità. Se prendiamo due volte la stessa decisione, non otteniamo necessariamente due volte la stessa domanda.

Aggiungiamo alcune ipotesi non strettamente necessarie e che possono essere rimosse, ma che semplificano il modello: *inizialmente non abbiamo informazioni sulla domanda, le domande dei prodotti sono indipendenti le une dalle altre e sono stazionarie nel tempo*.

L’ipotesi di assenza iniziale di informazione significa che alla prima decisione non siamo in grado di fare alcuna previsione sulla domanda, quindi non sappiamo né cosa aspettarci né come giudicare se un certo livello di domanda è buono o no. Scopriremo le reazioni del mercato alle nostre decisioni solo con l’esperienza della campagna di vendita.

L’ipotesi di indipendenza delle domande permette di ignorare fenomeni di elasticità incrociata, e quindi prodotti sostituti e complementari.

L’ipotesi della domanda stazionaria permette di ignorare diversi fenomeni:

- *Eccesso di domanda*: se un visitatore chiede di acquistare un prodotto e noi rifiutiamo la richiesta, ciò non modifica il comportamento futuro del visitatore, che nelle ore successive potrà decidere di comprare o non comprare per motivi indipendenti da quanto accaduto.
- *Saturazione della domanda*: il numero di compratori non varia in dipendenza dalle vendite delle ore precedenti. Questo non è vero se la popolazione dei visitatori è in numero finito e se ogni visitatore fa al più un solo acquisto. Stiamo quindi ipotizzando un mercato di dimensioni infinite, o quantomeno così grande da rendere l'effetto saturazione trascurabile.
- *Apprendimento dei compratori*: i visitatori non imparano nulla sulla nostra politica di vendita. Non sono quindi in grado di formulare previsioni sul nostro comportamento futuro e modificare le loro politiche di acquisto.

Le ipotesi di indipendenza e stazionarietà possono essere rimosse dal modello, e lo saranno in seguito.

3.3 Esempio: selezione del prodotto

Abbiamo tre prodotti, *A*, *B* e *C*. I prodotti hanno tutti lo stesso prezzo. Non ci sono vincoli di capacità. Ad ogni ciclo orario ci limitiamo a inviare al sito il nome del prodotto che vogliamo sia esposto. La risposta dell'ambiente esterno è il numero di unità vendute in un'ora, quindi la domanda d_t, che è funzione del prodotto scelto al tempo t. Per massimizzare il ricavo basterà quindi massimizzare le vendite.

Nella prima ora proponiamo un prodotto a caso, poniamo *A*, e dopo un'ora il sito ci dice che abbiamo venduto 3 unità. Dobbiamo scegliere il prodotto per la seconda ora. Non avendo informazione sulla domanda, non siamo in grado di sapere se 3 è un risultato buono o cattivo. Se avessimo mandato di realizzare un ricavo medio di almeno 2, allora avremmo un ragionevole motivo per insistere con la scelta di A, ma il nostro obiettivo è massimizzare il ricavo. La scelta naturale è *provare B*. Avvertiamo una chiara esigenza di saperne di più, di diminuire l'incertezza, anche se non è chiaro che cosa significhi qui incertezza. Vedremo che tale esigenza non è solo una sensazione di puro valore soggettivo, ma una necessità.

Scegliamo quindi *B* e dopo la seconda ora il server del sito ci invia le statistiche di vendita:

	Ore	Vendite	Media

A	1	3	3.00
B	1	1	1.00
C	0	0	-

La scelta per la terza ora pone un problema leggermente diverso. Adesso abbiamo dell'informazione di confronto fra prodotti: sappiamo che *A* è migliore di *B*, o meglio la stima di *A* è migliore. È chiaro che preferiamo *A* a *B*, ma è meno chiaro se preferire *A* a *C*. Potremmo pensare che *C* può essere il primo prodotto in redditività, il secondo o il terzo, mentre *A* può essere solo il primo o il secondo. Pertanto è "più prudente" preferire *A* a *C*. Però con questo principio non proveremo mai il terzo prodotto, e intuitivamente sentiamo che questo è scorretto. Scegliamo ugualmente *C* e a fine ora il server ci informa che ci sono state 4 vendite:

	Ore	Vendite	Media
A	1	3	3.00
B	1	1	1.00
C	1	4	4.00

Ora abbiamo provato tutti i tre prodotti e possiamo fare un confronto su base di parità. Il prodotto di maggiore performance economica, chiamiamolo il leader, è *C*, quindi lo scegliamo per la quarta ora. Si vendono 2 unità:

	Ore	Vendite	Media
A	1	3	3.00
B	1	1	1.00
C	2	6	3.00

Il leader *A* è preferibile a *B* per migliore media a parità di prove, e preferibile a *C* per minore campionamento a parità di media. Stiamo assegnando un valore al grado di campionamento: meno un prodotto è stato saggiato, più acquisisce un valore informativo che pesa a suo favore nella scelta.

Scegliamo *A* per la quarta ora e realizziamo 5 vendite:

	Ore	Vendite	Media
A	2	8	4.00
B	1	1	1.00
C	2	6	3.00

A è ancora certamente preferibile a *C*, ma nel confronto con *B* dobbiamo decidere se l'esigenza i saperne di più su B deve prevalere o no su una perdita attesa di 3 unità (la differenza fra le medie di *A* e *B*). Il confronto è su due criteri, performance e incertezza, e abbiamo bisogno di un modo per assegnare pesi ai due criteri in modo da renderli comparabili.

Dopo 15 ore la situazione è

	Ore	Vendite	Media
A	6	24	4.00
B	3	3	1.00
C	6	18	3.00

Le medie sono rimaste le stesse della quinta ora, ma ora il problema è diverso. Scegliere *B* è meno giustificato, sia perché dopo 3 prove abbiamo meno incertezza, sia perché mancando solo 9 ore alla fine diventa meno interessante imparare di più: se anche scoprissimo che *B* è il prodotto più redditizio, questo ci procurerebbe un vantaggio più ridotto rispetto a qualche ora fa. Dopo 23 ore, infatti, il problema dell'incertezza e dell'apprendimento non si pone più, e scegliamo semplicemente il prodotto più redditizio finora. Pertanto, il peso dell'apprendimento deve diminuire con il passare del tempo.

Riconduciamo questo esempio al nostro modello astratto.

Le azioni *A* sono semplicemente i prodotti:

Cosa siano gli stati *S* non è invece ovvio. Sembra anzi che non ce ne sia bisogno, perché per scegliere il prodotto per la prossima ora non è necessario sapere che cosa è accaduto nelle ore precedenti. Potremmo forse avere un modello con un solo stato, cosa perfettamente legittima. Ma riflettendo ci si accorge che nella discussione del nostro esempi abbiamo invece mantenuto la storia sotto forma di *informazione sulla performance delle decisioni precedenti*. Lo stato in questo modello è proprio la tabella ore-vendite, più formalmente è una lista di periodi in cui si è scelta l'azione e performance media dell'azione:

$$(ore\ di\ A, media\ di\ A), (ore\ di\ B, media\ di\ B), (ore\ di\ C, media\ di\ C)$$

Non che questa sia l'unica scelta possibile, ma è la più naturale.

Il tasso di sconto γ, come abbiamo detto, è 1 e quindi irrilevante.

La funzione σ è una trasformazione di tabelle di performance. Se l'azione scelta al tempo *t* è il prodotto *A*, lo stato s_{t+1} è ottenuto dallo stato s_t incrementando le ore di 1 e le vendite di r_t, il flusso generato dalla decisione, che è stata presa proprio basandosi sulla tabella s_t.

La funzione ρ dà in ogni ora il numero di vendite moltiplicato per il prezzo. Poiché il prezzo è uguale per tutti i prodotti, possiamo porlo a 1 e ignorarlo senza perdere in generalità, quindi il flusso coincide con il numero di vendite. Non c'è dipendenza dallo stato di arrivo

$$\rho(s, a, s') = \rho(s, a, s'')\ \forall s', s''$$

perché la domanda di un'ora non dipende dalle informazioni che il decisore avrà nell'ora successiva.

Una politica π è una qualsiasi regola che in base a una tabella ora-vendite sceglie il prodotto da promuovere nell'ora successiva. Se vogliamo che la regola cambi di periodo in periodo per tenere conto del tempo rimanente $T - t$, come abbiamo suggerito nell'esempio, allora lo stato deve essere esteso con questa informazione.

3.4 Esempio: selezione del prezzo

Abbiamo un solo prodotto e tre prezzi possibili, X, Y e Z. Non ci sono vincoli di capacità. Ad ogni ciclo orario inviamo al sito il prezzo che vogliamo sia esposto. Il ricavo orario è $p_t d_t$, il prodotto del prezzo esposto e della domanda, che è il numero di visitatori del sito che acquistano a quel prezzo. La domanda è funzione del prezzo: $d_t = d_t(p_t)$.

Le definizioni del modello sono simili a quelle vite per la selezione del prodotto.

Le azioni A sono i prezzi praticabili.

La definizione naturale degli stati S è la tabella ore-vendite, o meglio l'associazione fra i periodi in cui una azione è stata scelta e la sua performance media in quei periodi:

$$(periodi\ di\ X, ricavo\ medio\ di\ X), \ldots$$

Come nell'esempio precedente, possiamo inserire nello stato anche il tempo rimanente $T - t$, benché non sia necessario.

La funzione di transizione di stato σ è l'aggiornamento della tabella nella riga relativa al prodotto esposto, incrementando di uno il numero delle ore e ricalcolando il ricavo medio.

La funzione di generazione di flusso ρ è il ricavo dell'ora $p_t d_t$.

Le considerazioni svolte per la selezione del prodotto valgono ancora. Possiamo aggiungerne un'altra su una differenza sottile ma importante fra i due problemi della selezione del prodotto o del prezzo: i prezzi sono valori *numerici*, mentre i prodotti sono valori *nominali*. Tra valori numerici c'è un concetto naturale di *ordinamento*; se le azioni sono i prezzi 10, 20 e 40, sappiamo metterli in ordine e sappiamo che 20 è il valore intermedio. È naturale ipotizzare che la domanda sia funzione decrescente del prezzo, e in base a questa ipotesi noi acquisiamo *informazione sulle azioni non osservate*. Se nella prima ora esponiamo il prezzo 10 e osserviamo la domanda $d_1(10) = 7$, noi non soltanto sappiamo qualcosa in più sulla domanda del prezzo 10, e quindi sulla funzione $\rho(\cdot\,, 10, \cdot)$ ma anche su $\rho(\cdot\,, 20, \cdot)$, perché ci aspettiamo che sia maggiore.

Oltre al concetto di ordinamento, possiamo usare quello di *similarità* per acquisire informazioni sulle azioni non osservate. L'ordinamento ci dice che il prezzo 20 è intermedio fra i prezzi 10 e 40, e ipotizziamo che la sua domanda $d(20)$ sia intermedia fra le altre $d(10)$ e $d(40)$. La similarità ci dice che 20 è più simile a 10 che a 40, e si può ipotizzare che la sua domanda segua la stessa regola di prossimità. Questa nuova ipotesi appare però meno convincente della precedente, e questo ci mette in guardia dall'usare con troppa disinvoltura tale genere di supposizioni.

La similarità può essere usata anche per valori nominali: se riteniamo che i prodotti *A* e *B* siano più simili fra loro che con *C*, allora possiamo usare le informazioni su *A* e *C* per saperne di più (ipoteticamente) su *B*.

Queste considerazioni rientrano nel quadro della *generalizzazione dell'informazione*, la possibilità di usare l'informazione osservata sulle azioni per creare altra informazione ipotetica. È questa una possibilità reale, benché non semplice, e ne riparleremo. Per il momento, osserviamo che il nostro modello tratta le azioni come indipendenti, ipotesi semplificativa che in certi casi ci fa perdere interessanti opportunità. La generalizzazione dell'informazione può aggiungere molto al processo di ottimizzazione, ma di contro è complessa da trattare. Vedremo in seguito alcuni strumenti per arricchire in tal senso il nostro modello.

Nonostante queste limitazioni, trattare i prezzi come azioni nominali, senza tenere conto di ordinamento e similarità, è comunque possibile. Spesso i processi decisionali markoviani sono appunto affrontati in questo modo per limitare la complessità dei modelli e delle tecniche di ottimizzazione.

3.5 Esempio: selezione di prezzo e capacità

Abbiamo un solo prodotto e diversi prezzi possibili. La capacità del prodotto è limitata a 40 unità, quindi le vendite complessive non possono superare questa soglia. Ad ogni ciclo orario comunichiamo al sito sia il prezzo da esporre sia la capacità di periodo, quindi il numero massimo di richieste di acquisto da soddisfare. L'obiettivo è massimizzare il ricavo complessivo rispettando il vincolo di capacità.

A differenza ai due problemi di selezione del solo prodotto e del solo prezzo, qui abbiamo un problema di *ottimizzazione vincolata*: l'insieme della azioni varia secondo lo stato. Dobbiamo vedere come questo può essere rappresentato nel nostro modello.

Le azioni sono ora non più valori singoli ma coppie di valori prezzo-capacità, p.e. l'azione (10, 25) indica che nella prossima ora proponiamo un prezzo di 10 e siamo disposti a vendere fino a 25 unità.

Anche lo stato è più complesso, per l'esigenza di rappresentare sia l'informazione che la capacità. Abbiamo quindi una congiunzione di risorse materiali e immateriali; in generale, è utile pensare allo stato come un insieme di risorse. La capacità può essere inserita nello stato sotto forma di capacità residuale, l'invenduto fino a quel momento. Il tempo può essere inserito anch'esso in forma di risorsa residuale, il numero di ore ancora disponibili. Per quanto riguarda l'informazione, negli esempi precedenti abbiamo usato una tabella ore-vendite con una riga per ogni azione, m ora questa scelta non è più adeguata, per due motivi. In primo luogo, la risposta dell'ambiente all'azione selezionata non è il volume di vendite reale, ma la domanda, il volume di acquisti richiesti dai navigatori, quindi il volume di vendite che si sarebbe avuto in assenza di vincoli di capacità. In secondo luogo, la risposta dell'ambiente non dipende dall'azione prezzo-capacità, ma dal solo prezzo.

La soluzione che proponiamo è di definire lo stato come tabella prezzo-domanda:

Prezzo	Ore	Unità di domanda in media	Ricavo medio a capacità illimitata
10	5	42	420
20	8	28	560
40	7	16	640

La misura di performance non è il ricavo effettivo, che è $r_t = p_t \times \min\{d_t, c_t\}$, funzione della domanda ma anche della capacità di periodo. Misureremo piuttosto il ricavo in assenza di vincoli di capacità $r_t = p_t d_t$. La performance significativa sarà quindi la media di questo ricavo orario ipotetico.

Aggiungiamo poi allo stato il tempo residuo e la capacità residua, per esempio $T_{res} = 12$ periodi e $Cres = 75$ unità.

La funzione σ di transizione di stato consiste di tre operazioni:

- decrementare di 1 il tempo residuo;

- diminuire la capacità residua del numero di unità vendute nel periodo;
- aggiornare una riga della tabella prezzo-domanda.

Se abbiamo selezionato l'azione (prezzo = 10, capacità oraria = 25) e la domanda è stata di 30 unità, allora ne abbiamo vendute soltanto 25 ma nella performance ne includiamo 30.

Poniamo quindi $T_{res} = 12 - 1 = 11$, $C_{res} = 75 - 25 = 50$ e aggiorniamo la tabella

Prezzo	Ore	Unità di domanda in media	Ricavo medio a capacità illimitata
10	6	40	400
20	8	28	560
40	7	16	640

Abbiamo modificato la prima riga, relativa al prezzo selezionato, che non è l'azione ma solo al sua prima componente. Le 5 ore di selezione sono diventate 6. Le unità di domanda erano in precedenza $42 \times 5 = 210$. Ne aggiungiamo 30, la domanda non vincolata di questo periodo, e arriviamo a 240 unità richieste in 6 ore, con una media di 40 unità orarie, e un ricavo medio orario non vincolato di 400.

La funzione ρ di generazione del flusso non dipende dallo stato, ma dall'azione prezzo-capacità e dalla risposta dell'ambiente, cioè dalla domanda: $\rho(\cdot,(p_t c_t),\cdot) = p_t \min\{d_t, c_t\}$.

3.6 Selezione di prezzo e capacità con approvvigionamento

Ampliamo l'esempio precedente includendo la possibilità di approvvigionamento di unità di prodotto durante il processo decisionale. In ogni periodo possiamo acquisire una quantità illimitata di nuove unità di prodotto.

Le azioni sono ora triple di valori prezzo-capacità-acquisto, p.e. l'azione (10, 30, 18) indica che per la prossima ora esponiamo il prezzo 10, rendiamo disponibili per la vendita 30 unità di prodotto e ci approvvigioniamo di 18 nuove unità.

Lo stato rimane invariato rispetto all'esempio precedente e così la funzione di transizione stato.

La funzione di generazione di flusso cambia perché ora il ricavo deve essere al netto del costo delle nuove unità. Se il numero di unità acquistate è b_t e il costo unitario per l'approvvigionamento è *cb*, allora

$$\rho(\cdot, (p_t, c_t, b_t), \cdot) = p_t \times \min\{d_t, c_t\} - b_t \times cb$$

3.7 Considerazioni sul modello

Il modello presentato in questo capitolo trova la sua applicazione più naturale quando:

1. Il processo decisionale è markoviano, almeno approssimativamente. Intendiamo che un numero relativamente ristretto di variabili può mantenere la memoria degli aspetti più rilevanti per il decisore, che quindi non ha necessità di tenere costantemente conto di tutta la storia del fenomeno. Se così non è, la complessità del modello può diventare eccessiva.
2. L'apprendimento *in itinere* ha un ruolo importante nella decisione. Se il decisore ha già la capacità di prevedere in modo accurato la risposta dell'ambiente alle sue decisioni, e quindi non ha reale bisogno di apprendere durante il processo, allora il modello resta applicabile ma il concetto di stato che include l'informazione perde interesse.

Il primo aspetto, la proprietà markoviana, può sembrare prettamente tecnico, ma è essenziale per prendere decisioni razionali in molti contesti applicativi. Senza un modello più o meno markoviano, il decisore si trova ad affrontare un eccesso di informazione da filtrare, con un sovraccarico sia cognitivo per le persone sia elaborativo per i sistemi computazionali. L'indipendenza o almeno la debole dipendenza dalla storia del processo decisionale rende il problema più comprensibile e le soluzioni più realisticamente implementabili in sistemi software. Inoltre, e non è certo secondario, permette l'uso di tecniche matematiche e di apprendimento computazionale altrimenti inutilizzabili. Poiché questo modello non è semplicemente descrittivo ma piuttosto decisionale, non intende cioè solo rappresentare un fenomeno ma aiutare a guidarlo nel modo più efficace possibile, la moderazione della complessità è essenziale, anche se si deve pagare un inevitabile prezzo in termini di realismo descrittivo.

Il secondo aspetto, l'apprendimento, è il tema su cui il presente lavoro si incentra. L'idea di fondo è di rappresentare l'informazione come risorsa immateriale che gioca un ruolo essenziale nei processi decisionali volti al governo dell'equilibrio economico, e di riuscire a rappresentarla formalmente in modo logico-matematico. Vogliamo costruire modelli decisionali *autoregolati*, che contengono informazione su sé stessi e grazie a questa sono in gradi di migliorare le proprie performance in modo più o meno automatico.

Due dei prossimi capitoli saranno dedicati a due problemi cruciali che riguardano le potenzialità e i limiti dell'apprendimento nel nostro modello:

1. Il *dilemma dell'esplorazione*: vogliamo realizzare i migliori flussi possibili nell'immediato, ma al tempo stesso potenziare la nostra potenzialità futura di generare flussi; come bilanciare al meglio queste due esigenze tendenzialmente conflittuali?
2. La *generalizzazione dell'informazione*: impariamo dall'esperienza a valutare l'efficacia delle nostre decisioni, ma abbiamo anche convinzioni preesistenti al processo decisionale che esprimono informazione utile; come possiamo usarle per imparare dal'esperienza anche riguardo a ciò che non sperimentiamo direttamente?

3.8 Sul concetto di predizione

Nel quadro concettuale del nostro modello trova posto in modo naturale il concetto di *predizione*, che interpreteremo in modo non del tutto intuitivo. Vogliamo qui abbozzare una formalizzazione di questo concetto.

In primo luogo, la predizione consisterà nel generare *informazioni su transazioni future* basandosi sulle *informazioni su transazioni passate*. In questo enunciato la parola *informazione* assume due valenze diverse: le informazioni sul passato di regola sono oggettive, certe e precise, mentre quelle sul futuro dipendono da un modello matematico o algoritmico, e quindi sono soggettive, incerte e approssimate. La distinzione, peraltro, non è del tutto rigida, perché anche le informazioni sul passato dipendono dai metodi di rilevamento, e quindi possono essere in certa misura soggettive (perché si è scelto di rilevare certi dati e non altri), incerte (perché le informazioni sono incomplete o soggette a errore), approssimate (perché la rilevazione è approssimata).

Nella prospettiva del venditore ogni possibile futura transazione con il compratore avverrà in un contesto conoscitivo delineato dalle informazioni disponibili al venditore. Per esempio, un operatore della grande distribuzione sa distinguere tra vendite nella città *A* e nella città *B*, perché ha due punti vendita distinti. L'analisi degli scontrini permette di avere un database storico distinto per punto vendita e da questo trarre predizioni sulle future possibili vendite nei due punti. Quindi può fissare il futuro prezzo per *A* e il futuro prezzo per *B*. Se avesse un solo punto vendita non potrebbe farlo. Se gli ordini avvengono via Internet o telefono cellulare, e il sistema informatico del venditore è in grado di tracciare la provenienza geografica, la *capacità discriminante* dei dati è molto più elevata, e il venditore può praticare un gran numero di prezzi diversi.

Abbiamo detto che l'informazione sul futuro dipende da quella sul passato; aggiungiamo ora che dipende dal modello che usiamo per predire. Una cosa sono i dati su cui ci basiamo, un'altra gli algoritmi che usiamo per estrarne predizioni.

Il nostro particolare concetto di predizione è: *in base ai modelli e ai dati disponibili, delineare una relazione ipotetica tra i futuri contesti delle transazioni, le future offerte del compratore relative alle singole transazioni e le distribuzioni stocastiche delle risposte che i compratori daranno nelle singole transazioni*.

Un *contesto* di transazione è un assegnamento di valori alle variabili che descrivono ciò che noi (il venditore) sappiamo su ciò che crediamo influisca sul comportamento di acquisto. Esempi di tali variabili sono:

- Caratteristiche personali del compratore: età, genere, professione, interessi, tutte le classiche variabili di segmentazione dei clienti.
- Storia comportamentale del compratore: acquisti, rifiuti di acquisto, reclami, contatti di qualsiasi genere con noi o con altri soggetti rilevanti (p.e. concorrenti).
- Contesto temporale: data, orario.
- Contesto spaziale: stato, città, codice postale, punto vendita, indirizzo IP su web, coordinate rilevate da telefono cellulare, corsia del punto vendita rilevata via Bluetooth.

Un contesto potrebbe quindi essere: donna, fra 25 e 35 anni, con almeno 3 acquisti negli ultimi 6 mesi, di pomeriggio, in Italia. Gli esempi, come si vede, possono essere i più disparati: un contesto è ciò che sul nostro database accompagna una transazione, passata o futura.

I *dati* disponibili sono quelli che valorizzano i contesti di transazioni passate: il contesto appena descritto potrebbe essersi presentato poche o molte volte in passato, oppure mai. Oltre ai contesti, i dati disponibili contengono le *risposte*: nel contesto descritto, quante volte è stato venduto il tale modello di cellulare?

Le *offerte* sono ciò che noi proponiamo ai compratori, p.e. un certo modello di computer a un certo prezzo, a certe condizioni. A una offerta in un certo contesto segue una certa risposta del compratore.

I modelli di predizione sono "macchine" che prendono in input dati di contesto, offerta e risposta e restituiscono in output le distribuzioni stocastiche delle future risposte a certe offerte in certi contesti. Per esempio: se nel contesto su descritto offriremo il modello A di cellulare, la probabilità di acquisto sarà del 35%, per il modello B sarà il 60%. Un esempio più complesso di distribuzione di risposta potrebbe essere: per il modello A in quel contesto la risposta avrà una distribuzione normale con media 35% e varianza 20%.

Il decisore può avere a disposizione non un solo modello, ma molti, e in questo caso la predizione ha a che fare non solo con contesti, offerte e risposte che entrano in input nei modelli, ma anche con i modelli stessi fra cui scegliere.

Esprimiamo questi concetti in modo più formale.

$$Database\ predittivo = Contesti \rightarrow (Offerte \rightarrow Insiemi\ di\ risposte)$$

$$Modelli = Database\ predittivo \rightarrow Scenario$$

$$Scenario = Contesti \rightarrow (Offerte \rightarrow Distribuzioni\ di\ risposte)$$

$$Predizione = Insieme\ di\ modelli \rightarrow (Database\ predittivo \rightarrow Scenario)$$

I dati disponibili sono raccolti in un *database predittivo*. Il database predittivo è una funzione: preso in input un contesto restituisce in output una funzione, che a sua volta presa in input un'offerta restituisce in output un insieme di risposte realmente verificatesi in passato.

Un modello è una funzione: preso in input un database predittivo restituisce uno scenario.

Uno scenario è una funzione: preso in input un contesto restituisce una funzione, che a sua volta presa in input una offerta restituisce in output una distribuzione per ogni risposta.

Una predizione è una funzione: presi in input uno o più modelli restituisce una funzione, che preso in input un database predittivo restituisce uno scenario.

La differenza fra un database predittivo e uno scenario è che il primo contiene insiemi di risposte, dati oggettivi e certi (con le riserve di cui sopra), mentre il secondo contiene distribuzioni stocastiche di risposte, quindi oggetti ipotetici che sono soggettivi e incerti.

4 MODELLI E METODI DI DYNAMIC PRICING

In questo capitolo introduciamo informalmente il Dynamic Pricing e quindi definiamo dei modelli matematici per formularne alcuni tipici problemi. Daremo esempi di soluzioni con metodi euristici, di significato intuitivo per decisori che non siano esperti matematici e anche facilmente implementabili in software. In tal modo l'economista prenderà familiarità con l'operatività di questa disciplina.

Acquisita questa prima visione operativa, in seguito discuteremo più a fondo la logica economica e decisionale del Dynamic Pricing e il contributo imprescindibile che possono dare i metodi di Computational Learning e i moderni sistemi informativi.

4.1 Cos'è il Dynamic Pricing

Con la dizione di Dynamic Pricing si denota un vasto complesso di concetti e metodi per la gestione scientifica del prezzo di vendita a fini di ottimizzazione del reddito. Evidentemente il Dynamic Pricing è sempre esistito ed è connaturato al commercio, ma il trattamento scientifico delle politiche di prezzo richiede una disponibilità di strumenti matematici, dati e risorse computazionali che sono disponibili simultaneamente solo nell'epoca dell'informatica, e anche oggi non sono affatto scontate. I problemi concettuali e tecnici sono rilevanti, e non si possono ignorare le problematiche sociali, etiche e legali della gestione del prezzo.

L'esempio forse più familiare di Dynamic Pricing scientifico è oggi quello delle tariffe aeree per il trasporto passeggeri offerte dalle compagnie *low cost*. A differenza delle compagnie tradizionali, questi vettori non praticano la distinzione in classi di prenotazioni, ma tentano di ottimizzare i loro profitti usando la leva prezzo in modo assai accorto. Il punto critico è la non annullabilità della prenotazione[41], che impone a chi prenota il volo con largo anticipo un rilevante rischio di perdita del costo del biglietto in caso di cambiamento di piani. Ciò induce nel compratore una forte sensibilità al prezzo, alla quale la compagnia risponde con prezzi bassi. Con il trascorrere del tempo, il rischio del compratore diminuisce, e così anche la sua sensibilità al prezzo, consentendo al venditore di aumentare il prezzo stesso. Da qui il fenomeno dei prezzi crescenti, oggi familiare a

[41] Le prenotazioni sui voli low cost talvolta non sono annullabili, talvolta lo sono soltanto con penalità tali da rendere l'annullamento quasi equivalente alla perdita dell'importo già versato.

tutti. Non molti anni fa, invece, i prezzi dei voli erano molto più statici e la compagnie giocavano piuttosto sulla leva della capacità: suddividendo i posti in classi, influenzavano la domanda provocando scarsità nell'offerta. Si tratta di due strategie apparentemente opposte, in realtà due aspetti di una stessa filosofia: influenzare la domanda per massimizzare il profitto. Questa strategia, ancora oggi usata dalle compagnie tradizionali, è trattata altrove in questo lavoro.

Si dice comunemente che i prezzi dei voli *low cost* salgono all'avvicinarsi della data del volo. Non è del tutto esatto: di regola ciò che sale è la disponibilità a pagare dei compratori, e i prezzi sono regolati dai venditori per *tentare* di influenzare in modo ottimale la domanda. Si tratta di tentativi, non di manovre predeterminate e dall'esito certo. È infatti possibile che in certi momenti i prezzi scendano. Ciò può accadere per eventi esogeni alla dialettica prezzo-domanda, come scioperi o problemi metereologici, che di per sé riducono la domanda. Ma la riduzione di prezzo può accadere anche perché il venditore si accorge che la sua politica di prezzo per quel volo non ottiene l'effetto previsto dai suoi modelli predittivi. Se la *willingness-to-pay* dei compratori sta crescendo meno di quanto previsto, il venditore decide di ritardare la prevista crescita dei prezzi, rallentandola o anche facendola scendere temporaneamente per poi ripartire con gli aumenti. Può anche accadere il contrario: il venditore si accorge che la *willingness-to-pay* sta crescendo più velocemente del previsto e quindi accelera l'aumento dei prezzi[42].

I moderni sistemi informatici, con la loro disponibilità di dati in tempo reale e con sofisticati algoritmi che prendono decisioni in frazioni di secondo, non seguono più lo schema decisionale tradizionale:

1. analizza i dati storici e formula una predizione;
2. formula la strategia ottimale per quella predizione e seguila fino in fondo.

ma prediligono uno schema *adattivo*:

1. analizza i dati storici e formula una predizione iniziale;
2. formula la strategia ottimale per quella predizione e seguila per un breve periodo;
3. analizza i nuovi dati relativi al periodo e correggi la strategia per il prossimo periodo;
4. ripeti il ciclo.

[42] Quando diciamo che il venditore si accorge di un trend nella domanda e corregge la sua strategia, in pratica intendiamo che I suoi sistemi software correggono le predizioni e si adattano alla nuova forma della funzione domanda-prezzo.

Questo spiega l'estremo dinamismo dei prezzi dei voli *low cost*: ad ogni secondo nuovi dati suggeriscono nuovi prezzi, un po' come accade per le quotazioni di borsa.

L'esempio dei voli a basso costo non esaurisce affatto la gamma delle tipologie di Dynamic Pricing. Una logica diversa è all'opera nella distribuzione al dettaglio di abbigliamento, oppure di elettronica di consumo. Sono i tipici fenomeni dell'eliminazione delle giacenze e dei saldi di fine stagione, Qui i prezzi tendono a diminuire, anziché aumentare, con il passare del tempo.

Una spiegazione del comportamento dei compratori è che per loro il valore d'uso dei beni diminuisce con il tempo: il capo di vestiario non è più di stagione, o di moda, oppure il gadget elettronico non è più una novità, o un nuovo modello sta per uscire sul mercato. Nel caso dei voli non è così: l'utilità dell'acquisto di un biglietto per un volo non diminuisce all'approssimarsi della partenza, al contrario aumenta, perché aumenta la probabilità di fruire effettivamente del volo. L'acquisto del biglietto per un volo ha una componente di aleatorietà che manca nell'acquisto di un abito. In realtà non compriamo un volo, ma *l'opzione di fruire del* volo[43]. Possiamo aggiungere che l'uso di un abito o di nuovo computer ha una durata, e che l'acquisto tardivo implica una minore durata e quindi un minore godimento del bene, da cui la penalizzazione della disponibilità a spendere. Invece la fruizione di un volo non ha una vera durata, o meglio ha una durata indipendente dal tempo dell'acquisto. In definitiva, per il compratore un capo di vestiario ha una *deperibilità* che manca al volo, ed è intuitivo che i beni deperibili tendono a perdere valore con il tempo.

Fin qui le differenza fra i due casi dal punto di vista del compratore. Mettiamoci ora nell'ottica del venditore.

Nella distribuzione al dettaglio alla deperibilità del vestito per il compratore corrisponde un declino di valore per il venditore, perché diminuisce la *probabilità di vendere*: ecco che il venditore diminuisce il prezzo per far aumentare questa probabilità. Anche per la compagnia aerea il biglietto *è deperibile*: quando il volo parte il biglietto ancora invenduto non vale più niente, non perché il volo non è interessante, ma perché ormai la probabilità di venderlo in futuro è zero. La deperibilità per il venditore è di genere diverso da quella del compratore: ciò che deperisce è la probabilità di vendere. Perciò la compagnia aerea e il distributore al dettaglio hanno entrambi l'esigenza di

---

[43] La banale prenotazione di un biglietto aereo a basso costo implica quindi una qualche forma di *risk management*, sia pure semplice e quasi inconsapevole.

diminuire il prezzo nel tempo per contrastare la caduta della probabilità di vendere il bene (vestito o biglietto per il volo). La differenza è che per il distributore al dettaglio la sua tendenza a diminuire il prezzo e quella del compratore vanno nella stessa direzione. Invece, per la compagnia aerea c'è un conflitto fra la tendenza a diminuire il prezzo per diminuire il rischi di invenduto e quella ad aumentarlo per approfittare della crescente disponibilità a pagare dei viaggiatori. Entrambi vogliono eliminare le giacenze, ma per motivi e con meccanismi diversi.

I due tipi di venditori hanno anche un'altra esigenza in comune: estrarre informazioni dai loro clienti. Entrambi vogliono sapere qual è la propensione alla spesa dei clienti, per adeguare il prezzo e quindi appropriarsi del *surplus del compratore*[44] nella misura del possibile. Però, in questi due esempi, essi non sono in grado di praticare un prezzo differenziato per ciascun individuo, in parte perché non saprebbero come farlo, in parte perché tale comportamento sarebbe socialmente percepito come discriminatorio, e potrebbe anche avere conseguenze legali. Non potendo, o non volendo, operare una differenziazione di prezzo per individui, i due venditori ripiegano su una differenziazione di prezzo per segmenti di mercato. L'acquisto precoce o tardivo rivela, nel caso della distribuzione al dettaglio, che il consumatore acquista per motivazioni più legate allo stile di vita o alla economicità del bene; nel caso del trasporto aereo, se il viaggiatore ha più motivazioni di tipo *leisure* o di tipo *business*. I prezzi calanti servono a implementare una progressiva penetrazione in diversi segmenti dei rispettivi mercati, senza esplicitamente classificare i clienti in segmenti. I venditori fanno sì che i compratori operino una *auto-segmentazione*.

Questo spiega anche certi vincoli imposti dalle compagnie aeree, che altrimenti rimarrebbero incomprensibili. La migliore tariffa aerea per chi soggiorna il week-end fa si che i viaggiatori rivelino in qualche modo le loro motivazioni di viaggio. Chi rifiuta il vincolo di soggiorno rivela con ciò stesso che il suo viaggio è dovuto più a motivi di dovere che di piacere, il che suggerisce che probabilmente la sua sensibilità al prezzo è bassa[45]. Rivelando questo, il viaggiatore *business* assume una posizione negoziale più debole del viaggiatore *leisure*, che invece baratta volentieri un

[44] La differenza tra il prezzo massimo che il compratore era disposto a spendere e il prezzo che ha effettivamente speso. Dire che il venditore che si appropria del surplus del compratore significa che vende al prezzo massimo che il compratore può accettare.

[45] Chi viaggia per lavoro è più disponibile a spendere di chi viaggia per turismo, perché il suo tempo è percepito come di maggior valore economico, e anche perché chi decide il viaggio (il manager) non è in genere chi lo paga (l'azienda). Di contro, è più insofferente a vincoli sulle date, ha minore sicurezza di fruire davvero del volo che prenota, pianifica i voli con poco anticipo. Perciò tende a non volare *low cost*, oppure a farlo ma prenotando tardi.

vincolo con uno sconto; si trova così a ricevere un'offerta a prezzo più alto. In altri termini, lo sconto per chi accetta un vincolo può essere meglio compreso se lo si vede come un aggravio di prezzo per chi non lo accetta.

Emerge che la variazione di prezzo nel tempo svolge diverse funzioni:

1. estrarre dai clienti informazioni sula loro propensione alla spesa;
2. raggrupparli in segmenti;
3. adeguare il prezzo alla propensione alla spesa e catturare un parte maggiore di surplus del consumatore per il venditore;
4. coprire in tempi diversi la domanda di segmenti diversi;
5. controllare il rischio di invenduto.

Queste funzioni non sono affatto indipendenti: tutte rientrano nella formulazione generale di massimizzare il reddito pilotando in modo ottimale la domanda con manovre sul prezzo.

4.2 Problema introduttivo

Per introdurre in modo più rigoroso i primi concetti di Dynamic Pricing e la sua relazione con il Computational Learning utilizziamo un esempio semplice e intuitivo[46].

Il modello del problema decisionale si può così esprimere in termini informali[47]:

5. Vendiamo un singolo prodotto. Non ci sono competitori e prodotti sostitutivi.

6. La vendita si estende su due periodi.

---

[46] L'esempio è tratto da K.T. Tallury e G.J. Van Ryzin, *The theory and practice of Revenue Management*, Springer, New York, 2004.

[47] Il modello richiede anche alcune assunzioni matematiche intuitive e non essenziali per la comprensione della nostra trattazione. Le funzioni di domanda devono essere differenziabili con continuità, strettamente decrescenti (quindi invertibili), limitate superiormente, tendenti a zero al crescere del prezzo. La funzione di ricavo deve essere finita per tutti i prezzi ammissibili, massimizzabile da un prezzo ammissibile, con derivata parziale rispetto alla domanda strettamente decrescente.

Ponendo come obiettivo la massimizzazione del ricavo anziché del profitto stiamo assumendo che i costi siano fissi, quindi non rilevanti per il problema decisionale. Se i costi sono variabili ma dipendenti in modo semplice dalle quantità, allora possono essere agevolmente incorporati nel modello modificando i prezzi.

7. La domanda, cioè il comportamento dei clienti, varia fra i due periodi, ma non all'interno di un periodo. I clienti non anticipano nel primo periodo la variazione di prezzo del secondo. Abbiamo un modello predittivo della funzione di domanda.

8. Possiamo fissare liberamente i prezzi per i due periodi, ma non variarli durante un periodo.

9. La domanda dipende solo dal tempo e dal prezzo.

10. Il nostro obiettivo è massimizzare il ricavo totale nei due periodi.

Prima risolviamo il problema in termini di analisi matematica, poi passiamo a considerazioni più intuitive.

I prezzi dei due periodi siano p_1 e p_2, le domande d_1 e d_2. Il nostro modello predittivo della domanda è:

$$d_1 = 100 - p_1$$

$$d_2 = 120 - 2p_2$$

Nel secondo periodo i clienti sono più sensibili al prezzo: se fissiamo il prezzo a 30, vendiamo 70 unità nel primo periodo e 60 nel secondo. Il prezzo di saturazione, che azzera la domanda, è 100 nel primo periodo e 60 nel secondo.

La funzione da massimizzare è il ricavo totale:

$$r = r_1 + r_2$$

$$r_1 = p_1 d_1 = p_1(100 - p_1)$$

$$r_2 = p_2 d_2 = p_2(120 - 2p_2)$$

$$d_1 \geq 0, d_2 \geq 0$$

Per massimizzare il ricavo totale possiamo semplicemente massimizzare separatamente i due ricavi di periodo r_1 e r_2. Applichiamo il classico metodo dell'analisi matematica: imponiamo che la derivata prima della funzione ricavo sia nulla a e la seconda negativa:

$$100 - 2x = 0$$

$$-2 < 0$$

La seconda condizione è sempre soddisfatta, la prima lo è per p_1=50. Quindi nel primo periodo dobbiamo fissare un prezzo di 50; la domanda sarà 50 (secondo il nostro modello predittivo) e il ricavo 2500. Lo stesso procedimento ci dà il prezzo ottimale di 30 per il secondo periodo, con domanda pari a 60 e ricavo pari a 1800. Il ricavo totale ottimale sarà dunque 4300, con 110 unità di prodotto vendute.

Il problema diventa più interessante se disponiamo di meno di 110 unità da vendere, quindi se la nostra *capacità* è soggetta a vincoli. In questo caso non è più sufficiente ottimizzare separatamente i due ricavi di periodo, ma dobbiamo tenere conto di effetti inter-temporali tra i due periodi: una unità di domanda in più in un periodo è una unità di domanda potenziale disponibile in meno nell'altro periodo.

L'apparato matematico per la soluzione di questo problema è più complesso, ma i risultati sono del tutto intuitivi.

Supponiamo che la nostra capacità sia di C unità di prodotto. Non c'è possibilità di ri-approvvigionamento, quindi la somma delle domande in tutti i periodi non può superare C. Il problema di ottimizzazione si formalizza così:

$$\max \sum_{t=1}^{T} r(t, d_t)$$

$$\sum_{t=1}^{T} d_t \leq C$$

$$d(t) \geq 0$$

Indichiamo con $J_1(d_1)$ la derivata parziale del ricavo rispetto alla domanda nel primo periodo, e analogamente per $J_2(d_2)$. Indichiamo poi con d_1^* la domanda ottimale nel periodo 1, e analogamente per d_2^*; si tratta delle due grandezze incognite, proprio quelle che stiamo cercando di determinare per poter regolare in modo ottimale le due leve di prezzo. La teoria dell'ottimizzazione vincolata ci dice che esiste una terza grandezza π^*, detta *moltiplicatore di Lagrange*, per la quale valgono queste condizioni:

$$J_t(t, d_t^*) = \pi^*$$

$$\pi^*(C - \sum_t d_t^*) = 0$$

$$\pi^* \geq 0$$

Interpretiamo il significato economico di queste *condizioni di ottimalità*.

Il moltiplicatore di Lagrange π^* è il *costo opportunità marginale della capacità*, un “costo ombra” che dobbiamo sopportare quando vendiamo una unità supplementare di prodotto in un certo periodo, e con ciò rinunciamo all’opportunità di venderla in un altro periodo (o di non venderla). La seconda condizione di ottimalità dice che se fissiamo le domande ai livelli ottimali (incogniti), allora almeno una di queste condizioni deve essere soddisfatta:

1. Il costo ombra π^* *è* zero (e allora non è utile vendere di più).
2. La capacità rimanente $C - (\sum_{t=1}^{T} d_t^*)$ è zero (e allora non è possibile vendere di più)[48].

La prima condizione di ottimalità dice che, se le domande sono tutte ai livelli ottimali, allora i ricavi marginali di tutti i periodi sono uguali fra loro. Se così non fosse, allora sarebbe possibile manovrare i prezzi in modo da spostare una unità di domanda da un periodo a uno di maggiore ricavo marginale, ottenendo un ricavo più alto nel rispetto dei vincoli del problema.

Qui si mette a fuoco una attitudine mentale che ha un posto rilevantissimo nel repertorio concettuale di un decisore: il *criterio marginalistico*, su cui torneremo.

Ora, per dare concretezza al ragionamento matematico, torniamo al nostro esempio dei due periodi, fissando al capacità a livello di C = 40 unità di prodotto vendibile. Sostituendo i numeri dell’esempio nelle condizioni di ottimalità ed esprimendo il ricavo in funzione della domanda si ottiene:

---

[48] Si noti che non è richiesta la vendita di tutta la capacità: la soluzione ottimale può prevedere che restino delle scorte di invenduto.

$$max\left[d_1(100-d_1)+d_2(\frac{120-d_2}{2})\right]$$

$$d_1+d_2\leq 40$$

$$d_1\geq 0, d_2\geq 0$$

Le derivate parziali del ricavo rispetto alla domanda sono

$$J_1(d_1) = 100 – 2d_1$$

$$J_2(d_2) = 60 – d_2$$

Porre il moltiplicatore di Lagrange a zero, e quindi i ricavi marginali a zero, ci porta a richiedere d_1=50 e d_2=30, i prezzi che avevamo trovato quando non c'era il vincolo di capacità Ora queste domande non sono più ammissibili, perché richiederebbero la vendita di 80 unità di prodotto. Perciò nella seconda condizione di ottimalità dobbiamo imporre che sia a zero la capacità residua, e che quindi si vendano tutte le 40 unità disponibili. Perciò le condizioni di ottimalità diventano

$$100-2d_1=60-d_2$$

$$d_1+d_2=40$$

I due ricavi marginali sono uguali e la capacità è esaurita dalle domande.

La soluzione del sistema di due equazioni lineari in due incognite è

$$d_1=\frac{80}{3}$$

$$d_2=\frac{40}{3}$$

Approssimando la soluzione a numeri interi, per la natura del problema, la soluzione finale è

$$d_1=27$$

$$d_2 = 13$$

con prezzi $p_1 = (100 - 27) = 73$ e $p_2 = (120 - 13) \div 2 = 53{,}5$ e ricavo ottimale

$$r^* = 27 \times 73 + 13 \times 53{,}5 = 2666{,}5$$

4.3 Analisi della soluzione

Per comprendere intuitivamente i criterio di ottimalità dell'uguaglianza fra tutti i ricavi marginali, si può provare a spostare una unità di domanda da un periodo all'altro, lasciando il totale a 40 unità e regolando i prezzi in conseguenza.

Supponiamo di voler spostare una unità dal primo periodo al secondo, quindi con distribuzione 26 e 14 unità. Per ottenere questo risultato regoliamo i prezzi rispettivamente a 74 e 53, invece di 73 e 53,5. Il ricavo è allora

$$26 \times 74 + 14 \times 53 = 2666{,}0$$

Nel primo periodo abbiamo ottenuto un euro[49] in più sulle 26 unità di prodotto vendute, ma la 27-esima unità, quella "marginale", resta invenduta a causa della contrazione della domanda, e quindi perdiamo i 73 euro della vendita mancata, con saldo negativo di 47 euro. Nel secondo periodo vendiamo una unità in più a 53 euro, ma perdiamo 0,5 euro per ognuna delle 13 unità che avremmo comunque venduto, con un saldo positivo di 46,5 euro. Il saldo totale diminuisce quindi di 0,5 euro. Per inciso, nella soluzione 27 e 13 i ricavi marginali sono $J_1=46$ e $J_2=47$: sono quasi uguali, non perfettamente perché abbiamo arrotondato a numeri interi, e si vede che se fosse possibile spostare frazioni di unità di prodotto dovremmo spostare qualcosa verso il secondo periodo.. Nella soluzione 26 e 14 i ricavi marginali sono 48 e 46: sono più distanziati fra loro e ci dicono che dobbiamo spostare qualcosa verso il primo periodo.

Se viceversa spostiamo una unità dal secondo periodo al secondo, quindi con distribuzione 28 e 12 unità, otteniamo

$$28 \times (100 - 28) + 12 \times (120 - 12) \div 2 = 2664{,}0$$

[49] Usiamo l'euro come esempio di unità monetaria per alleggerire l'esposizione.

anche qui con una diminuzione di ricavo.

Risultati analoghi si ottengono se si diminuisce una delle due domande, portandole a 26 e 13 o 27 e 12 (aumentare non si può per il vincolo di capacità).

Questo controllo intuitivo è in sostanza un controllo sulle derivate parziali del ricavo, in accordo con la formulazione matematica della soluzione[50]. Ha però il pregio di mettere in chiara evidenza l'importanza del ragionamento marginalistico, che fa parte del repertorio concettuale del decisore di formazione economica e che può essere utilizzato spesso anche in mancanza di competenze o strumenti matematici adeguati.

Riformuliamo ora ciò che abbiamo fatto in un altro modo: abbiamo verificato che *a partire dalla soluzione* $d_1 = 27$ *e* $d_2 = 1$*, ogni piccola variazione ammissibile porta a un peggioramento del risultato*. Con la terminologia precedentemente introdotta, *abbiamo effettuato una ricerca locale a partire da una soluzione candidata e abbiamo verificato che si tratta di un massimo locale*. Non abbiamo in realtà verificato che si tratti anche di un massimo globale: non possiamo escludere che un'altra soluzione sia ugualmente "non migliorabile con piccoli passi" e migliore della coppia 27 e 13. Ma nella pratica questo problema teorico spesso non è rilevante: la soluzione ottimale è davvero unica. Nel nostro caso, la forma lineare delle funzioni di domanda lo suggerisce immediatamente.

Esaminiamo ancora la soluzione per trarne suggerimenti intuitivi.

Con alcune manipolazioni matematiche, la condizione di ottimalità sui ricavi marginali si può riformulare in termini di elasticità della domanda rispetto al prezzo:

$$\frac{p_t^* - \pi^*}{p_t^*} = \frac{1}{| \in_t \, (p_t^*)|}$$

dove

$$\in_t \, (p) = \frac{p}{d_t(p)} \frac{\partial d_t(p)}{\partial p}$$

[50] In realtà è un controllo sui rapporti incrementali, che diventano derivate solo con il passaggio al limite, ma il concetto resta valido.

Maggiore è l'elasticità in un periodo, minore è il prezzo ottimale di quel periodo. Se nei primi periodi i clienti mostrano maggiore sensibilità al prezzo che in seguito, allora i prezzi ottimali tenderanno a crescere nel tempo: questo è ciò che accade nel settore del trasporto aereo. Viceversa, se i clienti diventano più sensibili al prezzo al passare del tempo, allora i prezzi tendono a diminuire: ciò che accade nel dettaglio dell'abbigliamento.

Prendiamo il caso particolare in cui la funzione di domanda è la stessa in tutti i periodi. Allora il prezzo ottimale p_t^* è anch'esso uguale in tutti i periodi: quindi le variazioni di prezzo richieste dal modello dipendono solo dalle variazioni di domanda nel tempo. La seconda condizione di ottimalità richiede che almeno una fra due condizioni: i ricavi marginali sono tutti nulli oppure la capacità è esaurita. Chiamiamo *revenue maximizing price* il prezzo p^0 che rende uguali tutti i ricavi marginali, cioè $J_t(p^0) = 0$ per ogni t, e *stock clearing price* il prezzo p^C che permette la saturazione della capacità in T periodi, cioè $d(p^C)=C/T$. Allora il prezzo ottimo in tutti i periodi è

$$p^* = \max\{p^0, p^C\}$$

il massimo fra questi due prezzi. Il senso è che nessun prezzo può essere migliore di p^0; se p0 provoca un eccesso della domanda sulla capacità, allora il prezzo deve essere diminuito fino al livello p^C, il massimo sostenibile.

Il modello del problema e la sua soluzione mostrano quindi diverse proprietà in accordo con l'intuizione e l'esperienza comune.

4.4 Generalizzazione del problema

Il problema è stato presentato nella forma di differenziazione del prezzo in base al periodo. La stessa logica vale per la differenziazione rispetto al prodotto o al mercato, in generale per la differenziazione di prezzo rispetto alle combinazioni PMT.

Si possono i fatti sostituire i due periodi con due mercati, che possono essere due regioni geografiche come segmenti di clienti: la nostra definizione di mercato richiede soltanto che i clienti reagiscano diversamente agli stessi prezzi e che questa diversità di reazioni sia suscettibile di essere sfruttata per l'ottimizzazione del reddito.

Una interpretazione meno intuitiva sostituisce i due periodi con due prodotti, e il prodotto con un fattore produttivo. I due prodotti "pagano" il fattore che serve a entrambi, e che è disponibile in misura limitata (le 40 unità dell'esempio). Che cosa si vuole massimizzare, adesso? Non più il reddito in sé, ma qualcosa, poniamo la quantità di prodotto, che a sua volta influenzerà il reddito. Ma qual è ora il ruolo del prezzo? Il prezzo che il prodotto paga al fattore è certamente un concetto astratto, ma a ben vedere è un concetto base dell'economia aziendale: i fattori produttivi devono essere remunerati. La natura concreta di questa remunerazione dipende dalla natura dei fattori: costi dei fattori produttivi a rapido rigiro, quote di ammortamento di quelli a lento rigiro, oneri sui finanziamenti non di rischio, profitti per i finanziamenti di rischio. Questi *prezzi virtuali* permeano la vita dell'azienda, anche se non sempre sono esplicitati in questa forma. Più spesso, non si prendono in considerazione i prezzi virtuali, ma piuttosto i *ricavi virtuali* che remunerano i fattori, inglobando anche le quantità di fattore.

Dobbiamo quindi mantenere ben presente che ciò di cui stiamo parlando rientra sempre nel quadro concettuale delle discipline economico-aziendali, anche se lo svolgimento del ragionamento segue uno stile non usuale.

Possiamo ora riformulare il nostro problema secondo la logica fonti-impieghi, così utile per descrivere i fenomeni che formano l'oggetto dei sistemi informativi contabili.

Abbiamo due risorse A e B e due impieghi X e Y Vogliamo allocare le risorse agli impieghi in modo ottimale, massimizzando il reddito totale derivante dai due impieghi.

In una prima soluzione, decidiamo le quantità di risorse da assegnare agli impieghi, costruendo la matrice delle quantità

$$Q = \begin{bmatrix} q_{AX} & q_{AY} \\ q_{BX} & q_{BY} \end{bmatrix}$$

dove q_{AX} è la quantità di risorsa A assegnata all'impiego X. Le quantità inserite dovranno obbedire a vincoli di capacità, per esempio al loro somma non dovrà superare 40.

L'impiego X trasforma le quantità in reddito, secondo la logica del ciclo economico-finanziario, e si genera la matrice del reddito

$$R = \begin{bmatrix} r_{AX} & r_{AY} \\ r_{BX} & r_{BY} \end{bmatrix}$$

Il nostro problema è massimizzare la somma degli elementi della matrice del reddito, e la leva decisionale che abbiamo è la matrice delle quantità.

Una prima osservazione è che questa formulazione è semplificata: nella realtà non si generano quattro redditi da sommare, ma un unico reddito che emerge non solo dalle 4 quantità ma anche dalle loro interdipendenze, perché la logica del ciclo economico-finanziario non è così lineare. Ciò è assolutamente vero: il nostro modello è astratto, e scompone un unico sistema in sotto-sistemi per analizzarli singolarmente e ricomporre i loro funzionamenti in uno solo. Così facendo si perdono molte interdipendenze (anche se non tutte), e questo è il prezzo da pagare per usare un modello. Il beneficio di questa astrazione è che ci rende possibile l'uso dei modelli matematici e computazionali che abbiamo visto, e che nella prassi delle imprese possono apportare immensi benefici.

In questa prima formulazione il prezzo è scomparso, e al suo posto è rimasta la quantità. Ciò non è casuale, e deriva del dualismo insito nell'ottimizzazione dell'equazione del reddito: le leve decisionali sono sia i prezzi che le quantità. Nella prassi, i metodi di ottimizzazione usano sia i prezzi sia le quantità, ma in genere non con la stessa enfasi, e risultano più orientati verso gli uni o le altre.

Ampliamo ora il nostro modello reintroducendo il prezzo come leva di ottimizzazione.

Il nostro modello attuale presuppone che noi possiamo decidere le quantità di risorse che si allocano agli impieghi. Questo è di regola più o meno vero nei processi di produzione, molto meno in quelli di vendita. Il punto è che i processi di produzione sono interamente sotto controllo, quelli di vendita no. I prodotti non hanno obiettivi propri, i compratori sì. I prodotti si limitano a trasformare i loro input di quantità in output di quantità secondo regole prefissate e note all'impresa, i compratori trasformano input di prezzo in output di quantità secondo regole che l'impresa non conosce, ma può al più sforzarsi di prevedere. I compratori sono a loro volta degli ottimizzatori, e le grandezze che ottimizzano non sono le stesse che l'impresa vuole ottimizzare, e sono anzi parzialmente in conflitto con quelle. Chi compra contribuisce ad *aumentare* il reddito dell'impresa, ma il suo obiettivo non è *massimizzare* tale reddito. L'ottimizzazione di processi sotto controllo totale come la produzione è in sostanza un problema di teoria delle decisioni, in cui c'è un decisore ottimizzante e una *natura* che non decide e non ottimizza. Per i problemi sotto controllo soltanto parziale, come le vendite, si ha un problema di teoria dei giochi, in cui i decisori ottimizzanti sono due.

Il nostro modello rende meglio conto di tutto ciò se diciamo che l'impresa non costruisce la matrice delle quantità, bensì la matrice dei prezzi

$$R = \begin{bmatrix} p_{AX} & p_{AY} \\ p_{BX} & p_{BY} \end{bmatrix}$$

dove p_{AX} è il prezzo unitario con cui l'impiego X dovrà remunerare la risorsa A. Sarà poi un sistema esterno, quello dei clienti, che in base a una sua logica che noi non possiamo influenzare e non conosciamo, ma su cui possiamo formulare previsioni, genererà *in modo stocastico* la matrice delle quantità, cioè i valori della domanda. A quel punto, le matrici dei prezzi e delle quantità determinano la matrice del reddito. Perciò il fenomeno è di natura stocastica, non deterministica come, in genere, i fenomeni tecnico-produttivi.

Questo modello astratto permette di inserire il Dynamic Pricing in un quadro concettuale più ampio, in cui gli obiettivi e i parametri di ottimizzazione possono essere i più vari. Tenendo presente questo quadro, il decisore ha opportunità di usare i metodi di ottimizzazione in modi più creativi di quanto potrebbe fare focalizzandosi esclusivamente sulle applicazioni tradizionali.

4.5 Soluzione computazionale

Il problema di esempio può essere risolto per via analitica con le ben note tecniche di ottimizzazione non lineare. A noi interessa mostrare come esso possa essere risolto con una tecnica euristica, che richiede solo una matematica semplice, usando un algoritmo di ricerca locale. Si tratta di un esempio di algoritmo *greedy* (avido, goloso): prende in considerazione solo le alternative di brevissimo periodo, le più immediate, e sceglie la migliore senza curarsi delle conseguenze di più lungo termine. Ciononostante, è in grado di risolvere problemi come quello di esempio, posti certi requisiti matematici peraltro abbastanza realistici e soddisfatti in molti casi pratici.

Per illustrare passo per passo il procedimento scegliamo un esempio ancora più semplice del precedente[51].

Abbiamo di nuovo due periodi, in cui le funzioni prezzo-domanda sono così modellate:

$$d_1 = 10 - p_1$$

$$d_2 = \frac{1}{2}(10 - p_2)$$

I ricavi marginali rispetto alla domanda sono

$$J_1(d_1) = 10 - 2d_1$$

$$J_2(d_2) = 10 - 4d_2$$

Abbiamo una capacità *C=10*, quindi possiamo vendere fino a 8 lotti di prodotto, dove un lotto è una quantità convenzionale.

Vogliamo trovare una coppia di valori per le due domande che soddisfi il vincolo di capacità (la somma non supera 10) e le condizioni di ottimalità: ricavi marginali uguali e costo ombra nullo oppure capacità esaurita.

Partiamo dal presupposto che ci sia una sola soluzione di massimo locale: è una ipotesi ragionevole che possiamo sempre verificare a posteriori (e di fatto nell'esempio è vera). Il procedimento è semplice: partiamo con le due domande a zero e poi aggiungiamo un lotto alla volta a una delle due domande scegliendo quella che ci dà il miglior risultato immediato (ecco la logica *greedy*) cioè che ha il maggior ricavo marginale.

La tabella illustra tutti i passi del procedimento:

t	d_1	d_2	J_1	J_2	r

[51] Anche questo esempio è ispirato a uno di Talluri e Van Ryzyn, ma riformulato per estenderne la portata.

0	0	0	10	10	0
1	1	0	8	10	9
2	1	1	8	6	17
3	2	1	6	6	24
4	2	2	6	2	28
5	3	2	4	2	33
6	4	2	2	2	36
7	5	2	0	2	37
8	5	3	0	-2	37
9	6	3	-2	-2	36

Con zero lotti allocati, i ricavi marginali sono 10 per entrambi i periodi, come si ricava dalle equazioni di domanda. Possiamo scegliere arbitrariamente uno dei due periodi, e scegliamo il primo.

Ora abbiamo un lotto allocato (seconda riga della tabella, con $t=1$) e i ricavi marginali valgono 8 e 10. È accaduto che *avendo scelto una opzione abbiamo diminuito il suo valore relativo rispetto all'altra*. È all'opera il fenomeno della *utilità marginale decrescente*, ben noto nella microeconomia: più scegliamo un corso di azione, più questo apporta benefici sempre minori.

Adesso è il secondo periodo ad avere un ricavo marginale maggiore, quindi allochiamo ad esso il secondo lotto. Le due domande pianificate[52] diventano quindi 1 e 1. Adesso i ricavi marginali sono 8 e 6: come dicono le equazioni, ad ogni allocazione di un lotto supplementare il ricavo marginale di primo periodo diminuisce di una unità, quello di secondo periodo di 4 unità. È di nuovo preferibile il primo periodo, e allochiamo ad esso il terzo lotto.

I ricavi marginali sono di nuovo in parità, 6 a 6. Scegliamo di nuovo in modo arbitrario, stavolta il secondo periodo. Ora abbiamo due lotti per periodo e ricavi marginali di 6 e 2. Allochiamo il quinto lotto al primo periodo e poi il sesto con la stessa logica seguita finora.

La sesta riga della tabella mostra che abbiamo allocato 4 lotti al primo periodo e 2 al secondo. I ricavi marginali sono uguali fra loro e valgono 2, quindi la prima condizione di ottimalità è soddisfatta. Però la seconda non lo è, perché il costo opportunità marginale non è nullo (vale appunto 2) e la capacità non è esaurita. Quindi vogliamo allocare ancora lotti (perché il ricavo marginale è positivo) e possiamo farlo (perché c'è capacità residua, possiamo produrre altri due lotti, oppure abbiamo ancora due lotti in magazzino).

Allochiamo il settimo lotto in modo arbitrario, poniamo al primo periodo, con le domande che ora valgono 5 e 2. Poi allochiamo l'ottavo al secondo periodo, perché più vantaggioso, e arriviamo ai valori di domanda 5 e 3. Qui accade che un ricavo marginale, quello di secondo periodo, è negativo. Le condizioni di ottimalità non valgono ancora al passo 7 (5+2 i lotti allocati) e non valgono più. al passo 8 (allocati 5+3 lotti). Come interpretare questi risultati?

Con semplici calcoli si vede che sia al passo 7 sia al passo 8 il ricavo è 37: le due soluzioni sono equivalenti. Al passo precedente (con 4+2 lotti allocati) il ricavo era di 36, quindi è stato giusto

---

[52] La sequenza temporale riguarda la pianificazione delle vendite, non le vendite reali. Quando avremo scelto le domande ottimali e quindi i prezzi ottimali, del procedimento usato per calcolarli non rimarrà più traccia nell'operatività.

proseguire. È facile vedere che all'ottavo passo noi dovremmo allocare soltanto mezzo lotto al secondo periodo: così facendo avremmo 5+2,5 lotti allocati, con i due ricavi marginali azzerati, le condizioni di ottimalità soddisfatte e un ricavo di 37,5 che è appunto ottimale.

Se ha senso vendere mezzo lotto di prodotto, allora allochiamolo al secondo periodo. Se invece dobbiamo usare solo numeri interi, allora possiamo scegliere di fermarci al passo 7 o al passo 8, indifferentemente.

In ogni caso, non dobbiamo procedere al passo 9, che viola i vincoli di non negatività di π^*. Infatti con 6+3 lotti allocati il ricavo è 36, e proseguendo non può che scendere ancora.

Il problema è quindi risolto con un punto ottimale in 5+2,5 oppure con due punti ottimali in 5+2 e 5+3, secondo che siamo tenuti a usare solo lotti interi o anche frazioni. In ogni caso, *non* dobbiamo esaurire la capacità di vendita.

Se la capacità fosse stata soltanto di $C=6$ lotti, allora al sesto passo avremmo raggiunto entrambe le condizioni di ottimalità: ricavi marginali uguali e capacità esaurita. La soluzione ottima sarebbe quindi di domande 4 e 2. In questo caso, i ricavi marginali positivi indicano che sarebbe utile allocare ancora lotti, ma dobbiamo fermarci per il vincolo di capacità.

Una volta trovate le domande ottimali, dobbiamo soltanto regolare i prezzi in conseguenza. Se scegliamo di vendere $d_1=5$ e $d_2=2$ lotti, allora utilizzando le equazioni prezzo-domanda fisseremo $p_1=5$ e $p_2=6$ ottimizzando il ricavo totale dei due periodi pari a 37.

Questo algoritmo è estremamente facile da programmare e come strumenti matematici richiede solo il calcolo della derivata della funzione di domanda. È uno strumento pratico efficace e ha il pregio di basarsi sul ragionamento marginalistico, naturale per l'economista. La sua semplicità ha il vantaggio di focalizzare l'attenzione non tanto sul procedimento algoritmico risolutivo del problema di ottimizzazione, ma piuttosto sulla costruzione di un efficace modello predittivo della domanda, problema assai più congeniale per il decisore con formazione economica.

4.6 Modello con approvvigionamento

Nel nostro modello la capacità è prefissata, quindi non è ammessa la produzione di nuoe unità di prodotto durante al campagna di vendita, né il ripristino delle scorte. Introduciamo ora questa nuova possibilità.

Inseriamo nel modello queste nuove variabili:

x_t = quantità disponibile per la vendita all'inizio del periodo *t*;

y_t = quantità di riordino durante il periodo *t*;

c_t = costo unitario di riordino nel periodo *t*;

h_t = costo unitario di mantenimento delle scorte nel periodo *t*.

Il modello esteso è:

$$\max \sum_{t=1}^{T} r_t(d_t) - h_t x_t - c_t x y_t$$

$$x_t = x_{t-1} - d_t + y_t$$

$$d_t, x_t, y_t \geq 0$$

Assumendo per semplicità che sia $x_0=0$, quindi non ci sono scorte iniziali.

La variabile chiave per la soluzione è il costo per soddisfare la domanda nel periodo *t* con le scorte di un periodo *s* precedente a *t*:

$$\gamma_{st} = c_s + \sum_{k=s}^{t-1} h_k$$

Allora

$$\gamma_t^* = \min_{s \le t}\{\gamma_{st}\}$$

è il minimo costo necessario per approvvigionare il periodo t. Indichiamo con s_t^* l'indice che minimizza il lato destro dell'equazione, cioè il periodo antecedente a *t* che offre il minimo costo di approvvigionamento per sodisfare la domanda del periodo t. I due periodi possono anche coincidere, nel caso in cui è conveniente approvvigionarsi nel periodo t stesso, diciamo *just-in-time*.

Il criterio marginalistico di ottimalità è:

$$J_t(d_t^*) = \gamma_t^*$$

La soluzione ottimale dovrà avere i ricavi marginali di periodo uguali al loro minimo cosato di approvvigionamento per la domanda ottimale di periodo. Se così non fosse, potremmo migliorare la soluzione aumentando la domanda in un periodo a scapito di un altro, e quindi la soluzione trovata non sarebbe ottimale.

Abbiamo così caratterizzato le domande ottimali. Riguardo alla quantità ottimale di riordino in ciascun periodo *s*, la possiamo determinare come segue. considerano tutti i periodi *t* successivi a *s* per i quali *s* è proprio il periodo in cui è più conveniente approvvigionarsi per soddisfare la domanda ottimale al tempo *t*:

$$y_s^* = \sum_{t:s_t^*=s} d_t^*$$

Per quanto la complessità del modello sia cresciuta, resta possibile risolverlo in modo intuitivo usando ancora il criterio marginalistico.

È interessante osservare che, in questo modello esteso con l'approvvigionamento, anche se la domanda è stazionaria, cioè uguale per tutti i periodi, ciononostante il prezzo ottimale può variare fra i periodi a causa delle variazioni dei costi di approvvigionamento, che dipendono sia dai prezzi di acquisto che dagli effetti intertemporali che determinano le domande nei vari periodi. Si aggiunga che verosimilmente ci saranno interdipendenze tra prezzi e quantità sia sul lato degli acquisti sia su quello delle vendite. Davvero una illustrazione matematica esemplare del *Kreislauf* di costi e prezzi.

4.7 Modello ad approvvigionamento limitato

Estendiamo ulteriormente il nostro modello, non solo per renderlo più realistico ma anche per esemplificare più a fondo l'inaspettata versatilità ed efficacia che possono avere strategie computazionali relativamente semplici.

Supponiamo che le quantità di riordino non siano illimitate, vuoi per limitata capacità produttiva porpria o di fornitori, vuoi per vincoli legali, di trasporto, di magazzino:

$$y_t \leq b_t$$

Se i ricavi marginali di periodo sono decrescenti, allora il nuovo problema, alquanto più complesso del precedente, può essere risolto con un semplicissimo algoritmo *greedy* e con noto algoritmi disponibili anche in comuni software di calcolo.

Poniamoci dapprima il problema di minimizzare i costi di acquisto e mantenimento per l'approvvigionamento della capacità che servirà a soddisfare le domande dei vari periodi.

Il problema può essere espresso con variabili già note:

$$\min \sum_{t=1}^{T} h_t x_t + c_t y_t$$

$$x_t = x_{t-1} - d_t + y_t$$

$$y_t \leq b_t$$

$$x_t, y_y \geq 0$$

Questo è un problema di minimizzazione di costi di flussi, ben noto in letteratura. Non lo affronteremo qui, essendo il nostro scopo in questo lavoro trovare metodi relativamente semplici e di ragionevole efficacia. In realtà, questo tipo di problema può essere risolto usando software di facile reperibilità. Abbiamo impostato un semplice foglio di calcolo utilizzando Microsoft Excel e la sua funzione *Risolutore*.

t	x	y	h	c	d	b		
1	0		1	2		3		
2	0		1	2		3		
3	0		1	5		3		

Abbiamo inserito i valori di *h*, *c* e *b* per ciascuna di tre fasi. I valori di *x* sono calcolati come indicato nel problema: il valore di *x* precedente, più quello corrente di *y* meno quello corrente di *d*. Ora introduciamo una distribuzione della domanda:

t	x	y	h	c	d	b		
1	0		1	2	0	3		
2	0		1	2	0	3		
3	-1		1	5	1	3		

Ci sarà richiesta una sola unità di domanda al terzo periodo, nessuna prima. Qual è la strategia migliore per approvvigionarsi?

Si noti che non avendo ancora valorizzato le quantità di riordino *y*, nel terzo periodo la capacità è, correttamente, indicata come negativa dal foglio di calcolo.

Ora definiamo la funzione obiettivo, chiamiamo *g*:

t	x	y	h	c	d	b		
1	0		1	2	0	3		
2	0		1	2	0	3		
3	-1		1	5	1	3		
g								

Nella cella vicino a *g* (la settima cella della seconda colonna) inseriamo la funzione obiettivo, che è semplicemente la somma dei termini *hx* più la somma dei termini *cy*.

Quindi nel riquadro di input del *Risolutore* inseriamo i vincoli, diciamo che *g* deve essere minimizzato e che le variabili *y* sono quelle da valorizzare. Il Risolutore dà immediatamente la soluzione:

t	x	y	h	c	d	b		
1	0	0	1	2	0	3		
2	1	1	1	2	0	3		
3	0	0	1	5	1	3		
g	3							

La soluzione è di approvvigionarsi nel secondo periodo, $y_{2=1}$. La soluzione è convincente: approvvigionarsi nel primo periodo non offre vantaggi e costringe a sopportare il costo di mantenimento nel primo periodo; approvvigionarsi nel terzo periodo fa risparmiare il costo di mantenimento, ma impone un alto costo di riordino. La funzione *g* vale 3 perché si ha un costo di 2 per il riordino e un costo di 1 per il mantenimento.

Modifichiamo i costi *c* dei primi due periodi:

t	x	y	h	c	d	b		
1	1	1	1	1	0	3		
2	1	0	1	3	0	3		
3	0	0	1	5	1	3		
g	3							

Ora è conveniente approvvigionarsi nel primo periodo, e il costo complessivo è ancora 3.

Modifichiamo ancora i costi *c* e *h* e diventa ottimale approvvigionarsi nel terzo periodo, a un costo complessivo di 5.

t	x	y	h	c	d	b		
1	0	0	1	3	0	3		
2	0	0	2	4	0	3		
3	0	1	1	5	1	3		
g	5							

Adesso che sappiamo come risolvere il problema ausiliario del calcolo della funzione *g*, possiamo affrontare il problema principale: assegnare alle domande dei vari periodi i valori ottimali.

Il metodo *greedy* parte da un assegnamento di 0 a tutte le tre domande. Ci si chiede poi a quale delle tre sia più conveniente assegnare il primo lotto di prodotto. Si prova cosa accade se lo si assegna al primo periodo, creando la soluzione candidata $d_1=1, d_2=0, d_3=0$. Si calcola quindi il valore di

$$f = \sum_{t=1}^{T} r_t(d_t) - g$$

Poi si fa lo stesso per le soluzioni candidate (0, 1, 0) e (0, 0, 1). La funzione *f* è il reddito ottimale per una data configurazione di domande, perché dal ricavo (che dipende dalle domande e dai prezzi e quindi è fissato) si sottrae il minimo costo possibile (tale perché trovato appunto ottimizzando). Quella delle tre soluzioni candidate che migliora più delle altre il reddito viene scelta, e diventa il punto di partenza della successiva iterazione. Se, per esempio, la seconda soluzione è la migliore, si riparte dalla soluzione candidata (0, 1, 0) e si valutano le tre soluzioni candidate (1, 1, 0), (0, 2, 0) e (0, 1, 1). La migliore diventa la nuova soluzione provvisoria e si continua. Il procedimento si arresta quando ogni ulteriore passo peggiora il reddito, anziché migliorarlo.

La logica è quella tipica della ricerca locale: posizionati su una certa soluzione, si esplorano quelle raggiungibili in un piccolo passo, si sceglie la migliore, ci si sposta là e si ricomincia, finché non si raggiunge un ottimo locale, cioè un punto in cui ogni piccolo spostamento è dannoso.

4.8 Affidabilità e adattamento dei modelli predittivi

Nelle sezioni precedenti abbiamo elaborato modelli dei problemi e poi soluzioni che assumevano come date le funzioni di domanda, che sono create da modelli predittivi, la cui efficacia è in sé stessa un problema.

Un modo per rendere più efficaci i modelli predittivi è di renderli *stocastici*, cioè di esprimerli in termini di probabilità e statistica. Non si dirà allora che la domanda assume un certo livello in funzione del prezzo, ma che essa fluttua attorno a quel livello con una certa distribuzione stocastica, poniamo una curva normale di Gauss con una certa media e una certa varianza.

Il passaggio ai modelli stocastici comporta certamente una maggiore accuratezza nelle predizioni, ma anche un aumento considerevole di complessità matematica e maggiori difficoltà di interpretazione intuitiva. Tuttavia, la letteratura riporta casi in cui la formulazione deterministica di problemi come quelli che abbiamo qui illustrato riesce ad approssimare in nodo efficace la soluzione data da modelli stocastici[53], specie se le capacità sono grandi e i periodi di tempo numerosi. Quindi il decisore di impresa può ragionevolmente adottare modelli deterministici relativamente semplici e soluzioni computazionali accessibili come quelli visti e confidare in risultati non ottimali ma comunque affidabili e razionali.

In una strategia di Dynamic Pricing implementata seguendo modelli come quelli visti, si massimizza il reddito sotto il vincolo che la domanda media rispetti i vincoli di capacità. Questo compito di ottimizzazione viene svolto sia dai modelli deterministici che da quelli stocastici, e nella pratica sembra essere di per sé sufficiente per dare risultati apprezzabili. Ma i modelli stocastici svolgono anche un secondo compito di ottimizzazione, perché fanno fluttuare i prezzi in risposta alle fluttuazioni della domanda stessa intorno alla sua media; questo può loro un vantaggio sui modelli deterministici. C’è poi un terzo aspetto: i modelli stocastici di successo tendono ad essere inizialmente piuttosto aggressivi e proporre prezzi alti, per poi eventualmente diminuirli in seguito. Qui riappare il problema dell’esplorazione e dell’ottimismo. In questo contesto, esso si manifesta come tendenza a riservarsi delle *opzioni*: praticando prezzi alti ci si riserva di approfittare di una eventuale domanda più alta del previsto, oppure di cogliere “al volo” fluttuazioni sopra la media.

---

[53] Vedi G. Gallego e G.J. van Ryzin, Optimal dynamic pricing of inventories with stochastic demand over finite horizons, Management Science, 40:999-1020, 1994.

Stiamo parlando qui di una evidenza più aneddotica che sistematica, che però trova ampia conferma empirica in particolari settori, come la pubblicità online.

Comunque sia definito il modello predittivo, è importante essere in grado di apprendere online, cioè durante l'applicazione stessa delle strategie suggerite dal modello. Come visto in precedenza trattando l'apprendimento per rinforzo e la programmazione dinamica, qui si può seguire una strategia ricorsiva multi-stadio. Se abbiamo un problema di Dynamic Pricing a 10 fasi, con una capacità iniziale di 100 unità di vendita, dopo la prima fase, in cui abbiamo venduto 8 unità, rivediamo il nostro modello predittivo, ridefiniamo le nuove curve di domanda e riformuliamo un problema analogo a quello iniziale con 9 fasi e 92 unità di capacità. Questo atteggiamento adattivo è possibile particolarmente nelle applicazioni in cui i dati arrivano in tempo reale e i software sono in grado di riformulare i modelli altrettanto in tempo reale. Peraltro questa è una situazione abbastanza normale nella economia online e specialmente nella pubblicità.

5 MODELLI E METODI DI CAPACITY CONTROL

Dopo aver trattato il Dynamic Pricing e quindi l'uso della leva di prezzo per l'ottimizzazione del reddito, esaminiamo ora la leva di capacità. Anche in questo capitolo partiremo da una prima descrizione informale, per poi definire alcuni modelli matematici, con esempi di soluzioni mediante metodi euristici

In seguito discuteremo più a fondo la logica economica e decisionale del Capacity Control e le sue relazioni con il Computational Learning e i moderni sistemi informativi.

5.1 Cos'è il Capacity Control

Nel capitolo sul Dynamic Pricing abbiamo visto come le compagnie aeree low-cost usino la discriminazione di prezzo: con il passare del tempo il prezzo tende a salire, confidando nel fatto che i viaggiatori più sensibili al prezzo tendano a prenotare più presto mentre quelli con alta *willingness to pay* lo fanno con breve anticipo. L'ipotesi è che ci siano due gruppi di viaggiatori, tradizionalmente chiamati *leisure* e *business*, i primi sensibili al prezzo e disposti a rinunciare al viaggio oppure ad accettare vari vicoli e disagi, i secondi poco sensibili al prezzo[54], insofferenti ai vincoli, che pianificano con poco anticipo.

Manovrando la leva di prezzo, le compagnie che praticano il Dynamic Pricing tentano di regolare la domanda in modo da ottimizzare la sua distribuzione nel tempo. In alternativa, si può pensare di manovrare più direttamente la domanda fissando *limiti di capacità*: in un certo periodo di tempo è prefissato il numero di biglietti vendibili. Resta vero che il prezzo sale di periodo in periodo, ma ora la leva principale è la quantità, non il prezzo. Il risultato è per certi aspetti simile, anche se le differenze sono significative.

Chiamiamo Capacity Control il metodo di gestione della domanda facendo leva principalmente sulla capacità, intesa come limitazione artificialmente imposta alla domanda. Questo era il metodo usato un tempo dalle compagnie aeree, ed è ancora usato dalle compagnie tradizionali, non low cost.

---

[54] I viaggiatori *business* sono spesso dei dipendenti delle imprese, che sono poco sensibili al prezzo perché in realtà non lo pagano loro.

Per dare concretezza al discorso, partiamo da un caso semplice, tratto dal settore del trasporto aereo passeggeri, storicamente all'avanguardia nell'ottimizzazione del reddito. Abbiamo 100 biglietti disponibili su un volo: questa è la *capacità totale*. I biglietti sono venduti in un arco di due periodi, diciamo due mesi. I prezzi dei due periodi sono diversi, il secondo maggiore, sono *prefissati* e costituiscono quindi un input per il nostro problema di ottimizzazione. Fissiamo i due prezzi a 20 euro nel primo periodo e 80 euro nel secondo periodo. Vogliamo limitare le domande nei due periodi, imponendo un tetto a ciascuna. La scelta dei due limiti è libera, e vogliamo fissarla in modo da ottimizzare il reddito, dati i prezzi assegnati. Per semplificare, ottimizzeremo semplicemente il ricavo totale, assumendo che i costi incrementali siano trascurabili e che quindi il ricavo e il margine di contribuzione siano sostanzialmente equivalenti. Massimizzando il ricavo stiamo dunque massimizzando il reddito. Se i costi non sono trascurabili i modelli diventano un po' più complessi, ma la logica che descriveremo resta valida.

Sembra naturale fissare due capacità, poniamo 70 biglietti per il primo periodo e 30 per il secondo, ma una semplice riflessione dice che così facendo vincoleremmo inutilmente la domanda del secondo periodo. Se nel primo periodo vendiamo 60 biglietti, è evidente che nel secondo accetteremo di vendere fino a 40 biglietti, tutti i rimanenti, non solo 30. Il motivo è che il nostro prodotto è *deperibile*: alla partenza dell'aereo i biglietti invenduti valgono di colpo zero.

Pertanto, fissiamo soltanto la capacità del primo periodo, il cosiddetto *booking limit*. Nel secondo periodo la capacità sarà automaticamente fissata al numero di biglietti rimasti invenduti dopo il primo periodo. Abbiamo quindi un'unica leva decisionale, il booking limit, che possiamo fissare a piacimento da 0 a 100. Come posiamo determinare il suo valore ottimale?

La domanda del primo periodo è una funzione stocastica del prezzo, quindi dobbiamo ragionare in termini di scenari, predizioni, probabilità, pay-off e massimizzazione del valore atteso. Per scegliere il giusto valore del booking limit dobbiamo avere una stima della domanda, e in base a questa valutare le conseguenze delle nostre decisioni.

Il modo di ragionare è simile, ma non uguale, a quello visto nel Dynamic Pricing. In effetti, prezzo e capacità sono due leve decisionali alquanto diverse: la domanda è una funzione del prezzo, sia pure in senso stocastico, mentre non è una funzione della capacità, è soltanto limitata superiormente dalla capacità. Matematicamente parlando, il prezzo è un parametro più potente della capacità. A ben vedere, anche nel Capacity Control ottimizziamo usando il prezzo, perché i prezzi sono scelti liberamente. Il punto è che vengono prefissati e dopo restano come dati di input costanti, non più veri parametri.

Nel gergo del settore aereo si parla di classi di prenotazioni. Il termine *classe* suggerisce una differenza qualitativa dei posti offerti, ma non è necessariamente così. I posti possono essere del tutto indifferenziati dal punto di vista del servizio offerto. Quando è così, la differenza non è nel servizio, ma nella prenotazione stessa. Le classi di prenotazioni si distinguono in base al momento in cui avvengono, quindi siamo in un caso di auto-segmentazione dei clienti lungo la dimensione temporale. In questo contesto le classi coincidono con i periodi.

Diversamente, nella suddivisione in classi dei posti a teatro le classi di prenotazioni non sono definite dal periodo, ma dal prezzo. In questo caso si applicano i criteri già visti per il Dynamic Pricing, peraltro non lontani da quelli che stiamo per esaminare.

5.2 Il problema a due classi

Vogliamo ricavare il massimo reddito da un volo in cui abbiamo una capacità totale di $C=100$ posti, tutti identici. Vogliamo distinguere le prenotazioni in due classi, corrispondenti a due periodi. I prezzi del biglietto per le due classi sono prefissati come $p_2=20$ *e* $p_1=80$. Le prenotazioni del primo periodo avranno prezzo p_2, quelle del secondo p_1. Si noti la convenzione di usare coem indice il numero di classe, seguendo quindi l'ordine cronologico inverso: il periodo t_1 è l'ultimo, il periodo t_2 il penultimo e così via. Vogliamo scegliere il valore ottimale per il booking limit b.

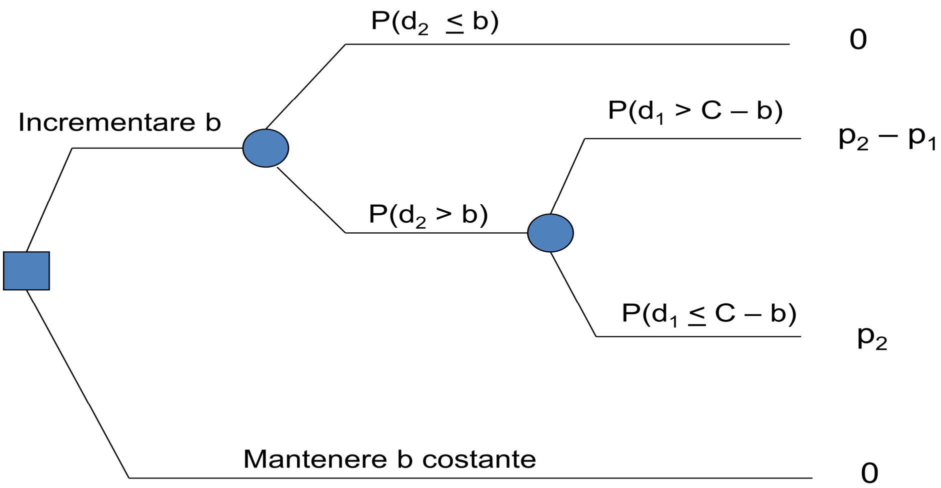
Abbiamo bisogno di un modello di previsione della domanda, che ci dica con quale probabilità la domanda può assumere un certo valore nei due periodi. Per esempio, il modello dovrà dici quanto vale $P(D_1>27)$, la probabilità che nel secondo periodo (prenotazioni di classe 1) la domanda ecceda le 27 unità. Assumiamo di avere disponibile un tale modello predittivo.

5.3 Regola di Littlewood

Per trovare il booking limit ottimale b^* usiamo il criterio marginalistico e la rappresentazione ad albero decisionale.

Ci chiediamo se aumentando *b* di una unità miglioriamo o peggioriamo il risultato, cioè il reddito: se abbiamo scelto $b=20$ è utile o no passare a $b=21$?

L'albero decisionale relativo alla decisione di incrementare *b* di una unità è[55]



Il valore addizionale atteso della decisone di mantenere *b* costante è *0*. Il valore addizionale atteso della decisone di incrementare *b* di una unità è

Se la domanda nel primo periodo non eccede il livello *b*, l'incremento non ha conseguenze. Se invece lo eccede, allora si hanno due conseguenze:

- è certo che si vende un biglietto in più per la seconda classe, con un flusso di reddito immediato nel primo periodo pari a p_2;
- è possibile che si venda un biglietto in meno per la prima classe, con una perdita differita al secondo periodo pari a p_1; ciò accade se nel secondo periodo la domanda eccede il numero di posti ancora disponibili, evento che ha probabilità $P(d_1 > C-b)$,

[55] Adattato da Phillips pg 150.

La decisione di incrementare *b* di una unità è preferibile se il valore addizionale atteso è positivo. La probabilità $P(d_2 > b)$ influisce sulla grandezza di questo valore, ma non sul suo segno, che è positivo se

$$p_2 - p_1 P(d_1 > C - b) > 0$$

cioè se

$$p_2 > p_1 P(d_1 > C - b)$$

o anche

$$P(d_1 > C - b) < \frac{p_2}{p_1}$$

Con i nostri valori numerici la condizione per decidere di incrementare il booking limit da 20 a 21 sarà

$$P(d_1 > 20) < 25\%$$

Nella loro semplicità, queste disequazioni facilitano interessanti intuizioni.

Introducendo la limitazione *b* abbiamo creato una scarsità artificiale per la seconda classe, e in questo modo abbiamo indotto una separazione fra due gruppi di prenotazioni. Si creano così dei costi opportunità. Accettando una prenotazione nel primo periodo, a fronte del *ricavo immediato e certo* p_2 c'è il *costo opportunità differito e incerto* $p_1 P(d_1 > C\text{-}b)$. Se il valore del booking limit è al livello ottimo incognito b^*, allora il ricavo e il costo opportunità marginali devono essere uguali. Se non lo sono, è possibile un arbitraggio, che aumentando o diminuendo il booking limit, sposta domanda dalla classe con minore ricavo marginale all'altra con incremento del risultato finale. Per essere precisi, si sposta domanda probabilistica e quindi ricavo atteso.

Possiamo anche dire che la vendita di un biglietto di seconda classe comporta un *rischio di rimpianto* se la domanda per la prima classe ci richiede quel biglietto marginale che però abbiamo già venduto a causa della decisione di incrementare b.

Se è poco probabile che la domanda per la prima classe arrivi al livello *C-b*, allora il rischio del rimpianto è piccolo, e anche un prezzo di seconda classe piccolo rispetto a quello di prima giustifica la decisone di accettare una prenotazione in più nel primo periodo. Al limite, se siamo certi che non ci sarà rimpianto allora siamo certi che è utile aumentare il booking limit. Pertanto, più si prevede piccola domanda nel secondo periodo, più è conveniente alzare il booking limit o accontentarsi un prezzo basso per la seconda classe. Parimenti, previsioni ottimistiche per il secondo periodo inducono a tenere basso il booking limit, oppure ad aumentare il prezzo per la seconda classe.

Nel nostro esempio numerico, un biglietto di prima classe vale 4 volte uno di seconda classe, quindi decido di vendere un biglietto in più nel primo periodo se prevedo che la probabilità di rimpiangere la decisione in futuro sarà minore di 1 su 4.

Il rapporto fra i due prezzi funge quindi da tasso di scambio fra le due classi di prenotazioni.

Vale anche la pena ricordare che le previsioni sulla domanda del primo periodo in realtà non influenzano il criterio per decidere se aumentare o no il booking limit. Davvero importante è avere buone previsioni sulla domanda del secondo periodo. Si ripresenta l'ordine cronologico inverso del processo decisionale. Per decidere, immaginiamo di soddisfare completamente la domanda del secondo periodo, per poi soddisfare quella del primo con i posti residuali. Tanto più accurate sono le previsioni della domanda di secondo periodo, tanto più possiamo avvicinarci ai risultati ideali che otterremmo se potessimo procedere a ritroso nel tempo.

Il livello ottimale b^* del booking limit deve rendere indifferente la scelta di allocare una unità addizionale di domanda fra i due periodi di tempo, e quindi soddisfa la condizione:

$$P(d_1 > C - b) = \frac{p_2}{p_1}$$

(si sceglierà il numero intero che meglio approssima la condizione).

Questa condizione è nota come *regola di Littlewood* e rappresenta uno dei primi risultati scientificamente formalizzati della disciplina del Capacity Control[56].

5.4 Soluzione computazionale euristica

L'analisi marginale suggerisce un semplice algoritmo per la computazione del livello ottimale del booking limit. Si usa una euristica *greedy*.

Si parte da $b=0$ e si incrementa b di una unità ogni volta che il valore addizionale atteso dalla decisione di incremento risulta positivo

$$p_2 - P(d_1 > C - b)p_1 > 0$$

Si arresta l'incremento quando questo valore diventa negativo o nullo, oppure quando il booking limit raggiunge la capacità. Il valore al momento dell'arresto è il booking limit ottimale.

La logica di questo algoritmo è la stessa dell'algoritmo che abbiamo usato per determinare le domande ottimali nel nostro esempio sul Dynamic Pricing.

Nel caso del problema a due classi, la regola di Littlewood permette di arrivare al risultato con un semplicissimo calcolo. Ma, come vedremo, la regola di Littlewood non si estende

[56] Kenneth Littlewood era un analista della BOAC, compagnia da cui nacque la British Airways. La sua regola di ottimalità fu enunciata nel 1972.

convenientemente al problema a molte classi, per il quale si deve comunque ricorrere ad un approccio algoritmico.

5.5 Euristiche EMSR

L'estensione della regola di Littlewood a più di due classi è possibile ma perde la semplicità che nel caso a due classi permette di evitare un approccio algoritmico.

Supponiamo di avere tre classi e di valutare se incrementare il booking limit b_3 del primo periodo[57], cioè il numero massimo di prenotazioni accettabili per la terza classe. Se nel primo periodo la domanda non supera b_3 allora l'incremento non ha alcun effetto. Se invece lo supera, allora l'incremento ci permette un ricavo addizionale immediato e certo di p_3, cui si contrappone una mancato ricavo differito e incerto, come già visto. Ma adesso c'è una complicazione: il mancato ricavo può manifestarsi nel secondo periodo ma anche nel terzo. Questo effetto spiazzamento (*displacement*), per cui una vendita in una classe può inibire una vendita in una classe superiore, è facile da trattare nel problema a due classi perché può nascere solo nel primo periodo e può manifestarsi solo nel secondo periodo. Ma nel problema a molte classi può nascere in un periodo qualsiasi, escluso l'ultimo, e manifestarsi in ogni periodo successivo. Non è semplice calcolare con quale probabilità la vendita incrementale nel primo periodo provocherà spiazzamento nel secondo, nel terzo o in entrambi. Entrano in gioco le probabilità condizionate: c'è spiazzamento nel terzo periodo a causa del primo se la domanda del terzo periodo è alta e se non c'è stato già spiazzamento nel secondo periodo a causa del primo[58].

---

[57] Ricordiamo che per convenzione l'indice si riferisce alla classe, quindi l'indice più alto è quello del primo periodo.

[58] Lo stesso biglietto venduto troppo precocemente non può essere rimpianto due volte, perché in ogni caso sarebbe stato venduto a prezzo migliore al più una volta.

L’estensione diretta della regola di Littlewood a molte classi è dunque possibile, ma implica la considerazione di ogni possibile sequenza di eventi che porta a un effetto spiazzamento, il che si traduce in una crescita esplosiva dell’albero di decisione. Pertanto non è utile in pratica.

Ciononostante, una versione euristica della regola di Littlewood è effettivamente usata nella pratica del Capacity Allocation. Si tratta dell’euristica *EMSR* (*Expected Marginal Seat Revenue*) introdotta negli anni ’80 da Belobaba[59].

Il procedimento è più chiaro se anziché di booking limit si parla di *protection level*: indichiamo con y_i il numero dei biglietti riservati alla classe i e a quelle superiori, quindi *protetti* dall’acquisto per classi inferiori. La classe più bassa, quella del primo periodo, ha protection level a 0. Se ci sono due classi allora c’è un solo protection level, $y_1 = C\text{-}b$, e i concetti di booking limit e protection level sono soltanto due modi complementari di vedere la stessa cosa.

La regola di Littlewood a due classi per il booking limit ottimale

$$P(d_1 > C - b^*) = \frac{p_2}{p_1}$$

si può riscrivere per definizione come regola per il protection level ottimale

$$P(d_1 > y^*) = \frac{p_2}{p_1}$$

Definiamo y_{21} come il livello di domanda per la classe 1 che soddisfa questa condizione:

[59] Riferimento.

$$P(d_1 \le y_{21}) = \frac{p_1 - p_2}{p_1}$$

Questa è ancora la regola di Littlewood per il problema a due classi. Nel nostro esempio avevamo detto che, essendo il prezzo di prima classe 4 volte maggiore di quello di seconda, la probabilità che la domanda superi il protection level $y^* = C - b^*$ deve essere il 25%. Ora diciamo che la probabilità di non superare il protection level deve essere il 75%, il che è lo stesso. Ma questa nuova formulazione è più agevole per rappresentare l'euristica EMSR.

L'idea è che in un problema a tre classi definiamo y_{32} come il protection level che soddisfa

$$P(d_2 \le y_{32}) = \frac{p_2 - p_3}{p_2}$$

e y_{31} come il protection level che soddisfa

$$P(d_1 \le y_{31}) = \frac{p_1 - p_3}{p_1}$$

Stiamo usando due volte la regola di Littlewood a due classi: una volta come se esistessero solo le classi 3 e 1, la seconda come se esistessero solo le classi 3 e 2: *stiamo ignorando le probabilità condizionate, come se gli effetti di spiazzamento tra periodi fossero tutti indipendenti*.

A questo punto il protection level per la classe 2 e superiori è semplicemente la somma dei due protection level già visti

$$y_2 = y_{32} + y_{31}$$

e il booking limit per la classe 3 è la capacità meno il protection level per le classi superiori

$$b_3 = C - y_3$$

Abbiamo quindi applicato ripetutamente la regola di Littlewood, ma sempre nella forma semplice a due classi.

Quella ora descritta è la prima forma dell'euristica EMSR, nota come EMSR-a; ne esiste una seconda forma, la EMSR-b. In questa versione, si assume che la vendita futura spiazzata da un incremento del booking limit di primo periodo abbia un valore che è dato dalla media dei prezzi delle classi superiori pesati con le rispettive domande attese. Si crea quindi una classe artificiale con tutte le classi superiori all'ultima e un prezzo atteso e si applica la regola di Littlewood al nuovo problema a due classi, l'ultima e la classe artificiale. Si calcola quindi il booking limit per la l'ultima classe. Poi si ripete il procedimento per la penultima classe e la classe artificiale che riunisce tutte le classi superiori, e così via. Anche EMSR-b è una applicazione ripetuta della regola di Littlewood sotto ipotesi semplificatrici.

Le due euristiche EMSR sono molto diffuse, sia perché ormai tradizionali, sia perché non richiedono sofisticate conoscenze matematiche ed algoritmiche, sia perché alcuni studi[60] indicano che i risultati ottenuti siano di poco inferiori a quelli ottimali ottenibili con i metodi di programmazione dinamica che stiamo per vedere.

5.6 Formulazione con la programmazione dinamica

Abbiamo $n > 2$ classi; nel primo periodo accettiamo le prenotazioni di classe n al prezzo minimo p_n, nel secondo periodo le prenotazioni di classe $n - 1$ al prezzo p_{n-1} e così via fino alla classe 1

---

[60] Riferimento Talluri, vedi Phillips p. 163

nell'ultimo periodo al prezzo massimo p_1. La formulazione in termini di programmazione dinamica richiede delle variabili di stato che racchiudano l'informazione necessaria per riformulare ricorsivamente il problema. Per noi la variabile di stato saranno x, la capacità rimanente dopo l'accettazione delle prenotazioni nei periodi precedenti, e D_j, la domanda che si manifesta in un certo periodo j.

Nel nostro modello, nel periodo j accadono questi eventi:

1. Osserviamo il manifestarsi della domanda D_j, che è una variabile stocastica.
2. Decidiamo il numero di prenotazioni u che accettiamo. Deve essere sia $u \leq min(Dj, x)$, perché non possiamo accettare più prenotazioni di quelle che ci vengono richieste, né più della capacità residuale di posti.
3. Realizziamo un ricavo $p_j u$.
4. Passiamo al periodo $j - 1$ con una capacità residua $x - u$.

Questa sequenza di eventi è ideale, non reale, e serve per facilitare l'analisi. Nella realtà noi non scegliamo la disponibilità u dopo avere osservato la domanda D_j; al contrario, scegliamo u a inizio periodo e poi riceviamo una alla volta le prenotazioni che costituiscono la manifestazione stocastica della domanda. Ma vedremo che la strategia per scegliere la variabile di controllo ottimale u^* non dipende dalla conoscenza a priori della domanda D_j, quindi possiamo procedere *come se* la domanda di periodo si manifestasse a inizio periodo, senza per questo perdere in generalità.

Indichiamo con $V_j(x)$ il *valore atteso nei periodi da j in poi assumendo che all'inizio del periodo j i posti ancora disponibili siano x*. Questo è un valore di stock, visto come potenzialità di futuri flussi cha andranno a costituire il ricavo totale nei periodi da j fino alla fine. È il valore atteso ottimale, sotto l'ipotesi che ad ogni periodo il problema sarà riformulato e risolto di nuovo secondo la strategia ottimale.

Dopo avere osservato D_j scegliamo il valore di *u in modo da ottimizzare la somma di valore di flusso immediato e valore di stock futuro*

$$p_j u + V_{j-1}(x-u)$$

rispettando il vincolo $0 \leq u \leq min\{D_j, x\}$.

L'equazione di Bellman è dunque

$$V_j(x) = \underset{D_j}{\mathrm{E}}\ [\max_{0 \leq u \leq \min\{D_j,x\}} \{p_j u + V_{j-1}(x-u)\}]$$

con condizione al contorno

$$V_0(x) = 0, x = 0, 1, \dots, C$$

dove $E[.]$ è l'operatore di valore atteso sulle possibili manifestazioni della domanda.

La condizione al contorno arresta il procedimento ricorsivo: il suo significato concreto è che dopo l'ultimo periodo, quando le prenotazioni sono chiuse e l'aereo è partito, il valore dei posti invenduti è 0.

La nostra *politica di controllo ottimale* consiste nel determinare ad ogni periodo il valore u_j^* che massimizza $V_j(x)$.

5.7 Soluzione ottimale marginalistica

Definiamo

$$\Delta V_j(x) \stackrel{\text{def}}{=} V_j(x) - V_j(x-1)$$

il *valore atteso marginale della capacità* nel periodo *j*, cioè il valore atteso incrementale della *x*-esima unità di capacità.

Si dimostra che il valore atteso marginale della capacità è *decrescente nella capacità e decrescente nel tempo*:

$$\Delta V_j(x+1) \leq \Delta V_j(x)$$

$$\Delta V_{j+1}(x) \leq \Delta V_j(x)$$

Sono queste le proprietà intuitive di un fenomeno a utilità marginale decrescente[61].

Decrescete nella capacità significa che avere 5 posti disponibili è meglio che averne 4, ma è "meno meglio" che averne 4 anziché 3. Il motivo è che per poter vendere il quinto posto, ammesso che vogliamo farlo, dobbiamo avere già venduto i primi 4 e avere l'occasione per vendere il quinto, due eventi che richiedono di avere già venduto i primi 3 e avere già avuto e sfruttato l'occasione di vendere il 4. Perciò il quinto posto è "più difficile da vendere" del quarto.

Decrescente nel tempo significa che avere un biglietto oggi è meglio che averlo domani. Il motivo è che oggi abbiamo un giorno in più per "manovrare" e trarne il massimo valore. Male che vada, questo biglietto disponibile oggi sarà usato come quello disponibile domani, sicuramente non peggio.

Il problema di ottimizzazione nel periodo *j + 1* può essere così formulato, dopo opportune elaborazioni algebriche:

$$V_{j+1}(x) = V_j(x) + \underset{D_j}{\mathrm{E}} \left[\max_{0 \leq u \leq \min\{D_{j+1}, x\}} \left\{ \sum_{z=1}^{u} \left(p_{j+1} - \Delta V_j(x+1-z)\right) \right\} \right]$$

[61] Stiamo usando il termine *decrescente* al posto del più rigoroso *non crescente*, ma è un piccolo abuso tradizionale.

dove la sommatoria vale 0 se $u=0$.

Traduciamo in parole l'equazione della politica di controllo ottimale.

Il valore di stock di oggi di x posti disponibili è uguale a quello di domani più il valore atteso delle vendite di oggi. Il valore atteso delle vendite di oggi è una media, pesata sulle probabilità degli scenari di domanda, dei valori che possiamo ottenere se scegliamo il miglior numero di vendite u per ogni scenario di domanda. Scelto u, ciascuno degli u biglietti venduti porta un ricavo pari al prezzo di oggi e un costo opportunità dovuto al fatto che domani avremo u posti disponibili in meno. I costi opportunità sono diversi per ciascun biglietto.

La variabile z serve come "cursore" per identificare il numero di sequenza dei biglietti. Sappiamo che il valore di stock è decrescente nella capacità, quindi il termine $p_{j+1} - \Delta V_j(x + 1 - z)$ è decrescente in z. Perciò è ottimale incrementare u e aggiungere nuovi termini finché si incontra un termine negativo oppure si raggiunge il limite superiore $min\{D_j, x\}$.

Il protection level y_j^* ottimale per la classe j e le superiori (numero di prenotazioni non accettabili per le classi sotto j) è

$$y_j^* \stackrel{\text{def}}{=} \max\{x: p_{j+1} < \Delta V_j(x)\}, 1 \leq j \leq n-1$$

con $y_n^* \stackrel{\text{def}}{=} C$ per convenzione.

Il controllo ottimale nel periodo $j + 1$è

$$u^*(j+1, x, D_{j+1}) = \min\left\{\left(x - y_j^*\right)^+, D_{j+1}\right\}$$

Indicando con $(.)^+$ la parte positiva di un numero: $(3)^+=3$ e $(-3)^+=0$.

La quantità $(x - y_j^*)^+$ è la capacità rimanente che eccede il protection level per i periodi da j (il prossimo) in poi, e quindi la massima capacità che vogliamo vendere alla classe $j + 1$ in questo periodo.

La traduzione operativa di questa politica ottimale di controllo nel periodo j + 1 consiste come primo passo nell'inserire il protection level y_j^* in un sistema di prenotazione; poi si accettano le richieste di prenotazione in ordine di arrivo, finché la soglia di capacità è raggiunta oppure il periodo finisce.

Come si vede, la politica di controllo ottimale del periodo corrente si basa sulle previsioni di domanda per i periodi successivi, ma non per il periodo corrente: u_{j+1}^* dipende da D_{j+1} ma non da D_j. Per questo era lecito immaginare di conoscere la domanda di un periodo già all'inizio del periodo, facilitando l'analisi senza ledere la generalità. Il motivo di questo è che la conoscenza della domanda presente D_{j+1} non influenza il valore futuro della capacità V_j. Per decidere se accettare o no una nuova richiesta abbiamo bisogno soltanto di confrontare il ricavo presente con il costo marginale della capacità, che dipende dalla domanda futura, non da quella presente.

Si può esprimere la politica di controllo ottimale in termini di booking limit anziché protection level, usando questa definizione

$$b_j^* \stackrel{\text{def}}{=} C - y_{j-1}^*, 2 \leq j \leq n$$

con $b_1^* \stackrel{\text{def}}{=} C$ per convenzione.

5.8 Bid pricing

Tradizionalmente si usano il booking limit e il protection level come controlli per l'accettazione delle prenotazioni. Esiste una interessante forma di controllo alternativa, basata non sulla classe ma sul ricavo, il *bid price*. L'idea è di definire una funzione prezzo-capacità $\pi(x)$ che dà una soglia di prezzo accettabile per ogni livello della capacità. Al diminuire della capacità residua la soglia di prezzo si abbassa. Una prenotazione viene accettata se ha un prezzo maggiore del bid price.

Il bid price può svolgere le stesse funzioni del booking limit e del protection level, con maggiore eleganza e semplicità formale. Ma soprattutto ha un grande vantaggio in termini applicativi: permette di trattare richieste con prezzi diversi nello stesso periodo. Si apre quindi la possibilità di discriminazioni di prezzo in una stessa classe. Il bid pricing può accettare solo le migliori proposte in una classe e rifiutare le altre, mentre i meccanismi di booking limit e protection level possono solo accettare o rifiutare in blocco tutte le richieste di una classe.

Il bid pricing, benché attraente, non è diffuso come i meccanismi orientati alla classe sia perché meno tradizionale, sia perché richiede sistemi informativi operanti in tempo reale, che sono davvero efficienti solo da non molti anni.

Nel problema a molti stadi la soluzione marginalistica trovata con la programmazione dinamica può essere espressa anche in termini di bid price ottimale.

Definiamo

$$\pi_{j+1}(x) \stackrel{\text{def}}{=} V_j(x)$$

Il bid price presente è il costo marginale futuro della capacità. La politica di controllo ottimale è allora:

$$u^*(j+1, x, D_{j+1}) = \begin{cases} 0 & se\ p_{j+1} < \pi_{j+1}(x) \\ \max\{z: p_{j+1} \geq \pi_{j+1}(x-z)\} & altrimenti \end{cases}$$

La richiesta numero z è accettata se il prezzo presente è maggiore del bid price, che è il costo marginale della z-esima unità di capacità.

5.9 Un metodo adattivo

Per determinare i parametri di controllo ottimi (booking limit, protection level o bid price) con i metodi descritti è naturale impostare un processo a tre fasi:

1. Esaminare i dati storici e costruire modelli della distribuzione della domanda.
2. Scegliere i parametri di questi modelli a scopo di previsione della domanda.
3. Usare le predizioni per trovare i parametri di controllo ottimi.

Il processo viene ripetuto periodicamente per aggiornare modelli, parametri e controlli. È un ciclo complesso, costoso e lento, e la lentezza può causare rilevanti costi opportunità dovuti al tardivo aggiornamento dei controlli al variare degli andamenti della domanda e dei vincoli esterni su prezzi e quantità.

In questa sezione esaminiamo come un metodo adattivo può essere usato per aggiornare *direttamente* i parametri di controllo del sistema di prenotazione, basandosi sull'osservazione della performance ottenuta con i parametri precedenti. In tal modo si può evitare, o almeno ridurre, il ricorso al ciclo descritto.

Il metodo ha alcuni requisiti non banali per garantire la convergenza alla politica di controllo ottima. Tuttavia, è rappresentativo di una classe di metodi che trovano spazio in molte applicazioni, specie nel mondo online.

L'idea alla base del metodo adattivo è questa: se i dati storici mostrano che i protection level attuali sono saturati troppo spesso dalla domanda, allora devono essere incrementati, se troppo poco spesso, allora devono essere decrementati. Se questi aggiustamenti hanno una giusta frequenza e intensità, possono mantenere i protection level bene adattati alla domanda.

Per definire il concreto criterio di "giusta frequenza di saturazione" ricorriamo a un importante risultato teorico. Denotiamo con B_j l'evento "la domanda eccede tutti i protection level della classi dalla j-esima in su". La probabilità di questo evento è funzione dell'insieme dei protection level e dalle domande delle varie classi, $P(B_j) = f(y_1, d_1, \dots, y_n, d_n)$.

Si dimostra che se tutti i protection level sono ottimali, allora per ogni classe j vale

$$P(B_j) = \frac{p_{j+1}}{p_1}$$

Intuitivamente, se così non è allora si possono proteggere di più alcune classi a scapito di altre. Si noti che se ci sono solo due classi, questa condizione di ottimalità si riduce alla regola di Littlewood.

Stimare la probabilità degli eventi di eccesso su dati storici è semplice, perché si richiede solo di osservare quante volte la domanda ha ecceduto i protection level in passato, non di quanto. Per esempio, se il prezzo p_3 è la metà del prezzo p_1, dovremo conteggiare quante volte in passato le domande per le classi 3, 2 e 1 hanno tutte superato i rispettivi protection level. Se questo è accaduto il 50% delle volte, allora y_3 ha superato il test di ottimalità.

La condizione di ottimalità suggerisce un criterio per l'aggiustamento dei protection level. Definiamo la quantità

$$h_j = \frac{p_{j+1}}{p_1} - P(B_j)$$

Nella soluzione ottimale h_j vale *0;* è negativa se l'evento di saturazione si verifica più spesso di quanto richiesto; è positiva se si verifica meno spesso. È allora intuitivo che se sottraiamo h_j da y_j l'aggiustamento va nella direzione corretta: a una saturazione troppo frequente si risponde con un incremento del protection level, a una saturazione troppo poco frequente con un decremento. Se siamo già nella situazione ottimale, i protection level restano invariati.

L'applicazione pratica consiste nell'aggiustare i protection level alla fine di ogni ciclo operativo, vale a dire ogni volta che si chiudono le prenotazioni e il volo parte. Con questo si chiude un episodio di apprendimento. Per ogni classe si verifica se c'è stato o no l'evento di eccesso di domanda per quella classe e tutte le superiori.

Se c'è stato l'eccesso delle domande, allora y_j deve essere incrementato e l'intensità dell'incremento deve essere forte se la probabilità attesa era bassa, debole se era alta. A un evento più inatteso si reagisce con una correzione più forte. Dobbiamo esprimere il "grado di sorpresa" dell'evento saturazione e per questo usiamo h_j, sostituendo la probabilità con il risultato empirico di periodo, quindi con *1* se l'evento si è verificato e con *0* se non si è verificato. Questo grado di sorpresa esprime l'intensità relativa della correzione. In definitiva l'aggiustamento è

$$y_j \leftarrow y_j - \left(\frac{p_{j+1}}{p_1} - 1\right) y_j$$

se l'eccesso di domanda si è verificato, altrimenti è

$$y_j \leftarrow y_j - \left(\frac{p_{j+1}}{p_1}\right) y_j$$

In realtà, per evitare oscillazioni continue e per convergere a risultati stabili grazie alla conoscenza accumulata nel tempo, dobbiamo anche in questo caso utilizzare un *tasso di apprendimento*, come già visto. Introduciamo il parametro γ e riscriviamo le due regole di aggiustamento così

$$y_j \leftarrow y_j - \gamma \left(\frac{p_{j+1}}{p_1} - 1 \right) y_j$$

$$y_j \leftarrow y_j - \gamma \left(\frac{p_{j+1}}{p_1} \right) y_j$$

Il tasso di apprendimento dovrebbe essere regolato in modo da decrescere con il tempo, se si suppone che la domanda sia stazionaria. Se non lo è, occorre garantire un certo tasso di apprendimento per catturare le variazioni[62].

[62] Non ci addentriamo nelel condizioni matematiche che garantiscono la convergenza verso la soluzione ottimale nel caso di domanda stazionaria. Si veda Talluri p. 52.

6 ECONOMIA DEL REVENUE MANAGEMENT

In questo capitolo esaminiamo i processi di formazione del prezzo nella prassi aziendale alla luce di quanto appreso circa le possibilità di manovra sui prezzi stessi offerte dai modelli di Dynamic Pricing. Assumeremo che l'impresa agisca in condizioni di monopolio o abbia in ogni caso buoni spazi di manovra per fissare i prezzi.

Estenderemo poi l'analisi al caso più ampio del Revenue Management, inteso come combinazione di Dynamic Pricing e Capacity Control, in cui si agisce anche sulla leva quantità, oltre che sul prezzo

6.1 Dimensioni delle decisioni di prezzo

Le decisioni sul prezzo tengono conto di tre dimensioni fondamentali:

1. *Economia interna*. L'azienda valuta i propri costi e investimenti, e il ruolo a loro assegnato nell'equilibrio economico, tipicamente nel conseguimento degli obiettivi di reddito.
2. *Concorrenza*. L'azienda valuta le politiche di prezzo dei concorrenti, formula previsioni sui loro sviluppi futuri, ipotizza le loro risposte alle proprie manovre di prezzo.
3. *Domanda*. L'azienda valuta le risposte dei clienti alle proprie manovre di prezzo e prevede i conseguenti andamenti di vendita.

L'azienda effettuerà quindi una analisi economico-organizzativa, una analisi competitiva e una analisi della domanda. Nella prassi, non necessariamente le decisioni si baseranno su tutte le tre dimensioni: spesso l'azienda userà una delle tre tipologie di analisi, riservando alle altre due un ruolo secondario. Avremo dunque tre tipologie di orientamento nelle decisioni di prezzo:

1. Orientamento ai costi.
2. Orientamento alla concorrenza.
3. Orientamento alla domanda.

I tre orientamenti sono distinti ma complementari, come illustrato dal ben noto modello di Monroe[63]:

[63] K. Monroe, Pricing. Making Profitable Decisions, McGraw-Hill, New York, 2003

L’analisi economica interna è utile a fissare un limite inferiore per il prezzo, sotto il quale viene a mancare l’equilibrio economico. L’analisi della domanda ci dà invece un limite superiore, oltre il quale il prezzo diventa non realistico perché superiore al valore per i clienti. Entrambe, insieme con l’analisi della concorrenza, ci aiutano a capire come muoversi nell’area discrezionale tra questi valori minimo e massimo.

6.2 Orientamento ai costi

L’azienda che segue l’orientamento ai costi intende assegnare al prezzo il compito di realizzare le condizioni per il mantenimento di un “adeguato” equilibrio economico. Si richiede una analisi dei costi, degli investimenti, degli andamenti finanziari e degli obiettivi di redditività. Tipicamente il prezzo dovrà coprire i costi diretti di prodotto, i costi indiretti e il reddito atteso. Più specificamente, si possono distinguere quattro metodi[64]:

a) *Mark-up*. Al costo diretto di prodotto si aggiunge un ricarico percentuale ritenuto idoneo a coprire una quota di costi indiretti e una quota del reddito atteso. È questo un metodo assai diffuso, gradito alle imprese che usano la contabilità a costi diretti. Appare semplice e intuitivo e quindi “rassicurante”, ma non è certamente un metodo rigoroso e presenta una evidente rigidità di applicazione. Si basa su regole tradizionali spesso semplicistiche, come la regola vigente qualche anno fa nel settore della ristorazione americana: il prezzo del cibo copre 3 volte i costi diretti, per la birra 4 volte, per i liquori 6 volte[65]. Dal nostro punto di vista, è palesemente inadatto a manovre di ottimizzazione, per la sua mancanza di flessibilità e perché per natura ignora proprio l’elemento cardine dei nostri modelli, la domanda, quello che ci dà gli obiettivi a cui tendere con l’ottimizzazione.

[64] Si veda Busacca, Costabile, Ancarani.

[65] Citazioen da Godin, vedi Phillips p.37.

b) *Full cost plus*. Le imprese che hanno una contabilità a costi pieni possono usare questa versione più avanzata del mark-up: il costo base include sia i costi diretti che indiretti, e il ricarico deve coprire una quota del reddito atteso. Vale quanto detto per il mark-up: si tratta di un metodo utile per fissare limitazioni inferiori al prezzo, ma inadeguato per l'ottimizzazione.

c) *Target price*. Il metodo si basa sulla *break even analysis*: il prezzo deve coprire tutti i costi più un livello atteso di reddito. La differenza rispetto al metodo full cost plus consiste nell'uso di simulazioni *what-if* per la ricerca del punto di pareggio. Anche se il punto di arrivo è pur sempre prezzo = costo + ricarico, tuttavia il metodo forza il management ad una analisi più accurata dei possibili scenari futuri e delle loro conseguenze.

d) *Marginal cost*. Il metodo consiste nell'applicare il mark-up ai soli costi marginali di produzione. Si tratta di un metodo applicabile in particolari condizioni. Se l'impresa copre già i costi fissi, questi vengono visti come *sunk costs*, che possono ormai essere ignorati. Se la capacità produttiva non è saturata, ha senso aumentare la produzione per venderla al costo marginale più un ricarico, anche molto basso. C'è il rischio di provocare danni di immagine o opportunità di arbitraggio.

Per i nostri interessi l'orientamento ai costi non è un metodo adeguato di formazione del prezzo, al più servendo come analisi del limite minimo accettabile. Un suo difetto evidente è la mancanza di flessibilità: incoraggia a "spalmare" in modo indifferenziato i costi sui vari prodotti, mercati e periodi. Un altro è la forte carica di soggettività: le valutazioni di costo hanno un ampio margine di discrezionalità. In sintesi, ignorando la domanda e quindi il valore per il cliente, con questo metodo si rinuncia in partenza a tentare di raggiungere il limite superiore del prezzo, quindi si rinuncia all'ottimizzazione stessa.

Contrario al nostro quadro concettuale è poi la stessa formulazione del problema in termini matematici: si vede il prezzo come funzione del reddito, mentre per noi è il contrario.

L'orientamento ai costi soffre di una disapprovazione teorica generale, ma non per questo è scomparso dalla prassi aziendale, conservando pur sempre una grande attrattiva psicologica e, cosa non da poco, richiedendo un modesto sforzo analitico e decisionale in quanto basato su dati già disponibili in azienda per scopi diversi dal pricing.

6.3 Orientamento alla concorrenza

L'impresa che segue l'orientamento alla concorrenza raccoglie informazioni sui prezzi praticati dai suoi competitori per prodotti simili ai propri e decide il proprio posizionamento di prezzo in riferimento a quelli. Possiamo distinguere diverse varianti di questo orientamento:

a) *Going rate pricing*. L'impresa pratica lo stesso prezzo dei suoi concorrenti. L'intero settore mantiene quindi lo *status quo*.

b) *Price taker – price maker*. L'impresa adegua costantemente il proprio prezzo per mantenere costante il differenziale con una o poche aziende di riferimento. Anche in questo caso il settore mantiene lo *status quo*, nel senso che i prezzi non sono tutti uguali ma i rispettivi posizionamenti non cambiano.

c) *Discount price* e *premium price*. L'impresa sceglie il propri posizionamento sul mercato e colloca il proprio prezzo, secondo i casi, sotto o sopra la media di mercato. Assomiglia al metodo del differenziale costante, ma è una decisione autonoma dell'impresa, secondo una logica competitiva non legata al mantenimento dello *status quo*.

L'orientamento alla concorrenza richiede non solo *l'intelligence* dei prezzi attuali dei concorrenti, ma anche delle loro serie storiche, per capire i trend. Impegna inoltre nella formulazione di previsioni e di strategie competitive. L'uso del *pricing* come strumento competitivo è un argomento affascinante, che però esula dagli scopi del nostro lavoro. Per i nostri interessi, il *pricing* orientato alla concorrenza ha un interesse accessorio rispetto a quel orientato ala domanda.

L'orientamento alla concorrenza è tipico dei mercati di *commodity*, nei quali i prodotti sono facilmente confrontabili e i prezzi delle transazioni rapidamente diffusi. In questi contesti vale il principio che in un mercato perfetto in realtà non ci sono vere decisioni di prezzo: un'impresa non è in grado di influenzare sensibilmente il prezzo di un prodotto con proprie decisioni.

Il metodo è anche utilizzato dai nuovi entranti in un mercato, che almeno inizialmente si adeguano ai prezzi correnti, e dai piccoli attori in un mercato dominato da grandi imprese, che svolgono obiettivamente il ruolo di *price-maker*. In questi contesti praticare prezzi diversi può essere molto rischioso, perché i clienti non percepiscono la ragione del prezzo differenziato e non sono motivati a indagare.

In altri contesti l'orientamento al mercato non è oggettivante richiesto dalle circostanze e appare come una facile scorciatoia, un abdicare dell'impresa alla propria facoltà di decidere il prezzo[66].

Anche l'orientamento alla concorrenza non appare adeguato ai nostri scopi, perché per definizione ignora il limite superiore dell'ottimizzazione, il valore per il cliente. Tuttavia, rispetto all'orientamento ai costi, è più probabile che ci possa dare utili indicazioni sulle possibilità di manovra all'interno dell'area discrezionale del prezzo.

---

[66] Abdicazione che però può trovare giustificazione nell'acquisizione a basso costo di conoscenza: imitare il comportamento di imprese simili significa imparare più rapidamente ed economicamente. Dopotutto, imitare un competitore che ottimizza i suoi prezzi si avvicina a ottimizzare i nostri prezzi. La condizione è che i competitori di riferimento siano davvero simili a noi, altrimenti perdiamo le opportunità offerte dalle nostre peculiarità, ad esempio una migliore percezione che il mercato ha di noi.

6.4 Orientamento alla domanda

L'impresa che segue l'orientamento alla domanda si sforza di misurare e prevedere il valore che i clienti attribuiscono al suo prodotto. Nella misura in cui ci riesce, essa si procura un formidabile strumento di ottimizzazione della redditività, perché conosce il limite massimo a cui può spingere il prezzo. In tal modo essa può appropriarsi di un'alta quota del *customer surplus*[67].

Il grande problema di questo orientamento è la misurazione del valore per il cliente. Esiste un'ampia gamma di tecniche, ampiamente trattate nella letteratura[68]. Ci limitiamo qui a segnalare alcune fra le più diffuse[69]:

- stima dell'elasticità della domanda;
- metodo *buy-response*;
- analisi EVC (*Economic Value for the Customer*);
- approccio alla Fishbein;
- *conjoint analysis*.

Negli ultimi anni, con l'avvento di Internet e di metodi computazionali avanzati, come il *data mining*, ha assunto un profilo più realistico l'idea di praticare il prezzo personalizzato, diverso per ciascun cliente in ciascun momento. Tale prezzo personalizzato intende approssimare il valore per *questo* cliente *qui e ora* con il prezzo. Indubbiamente il concetto è affascinante e di grande utilità come idea-guida. Occorre però ricordare che si tratta appunto di un ideale, non di un obiettivo da

[67] Non interamente, altrimenti il compratore non avrebbe più interesse all'acquisto dei nostri prodotti. Parliamo infatti di limite massimo, da approssimare per difetto.

[68] Citare Dalli e tutti gli altri.

[69] Rimandare a Busacca, Costabile, Ancarani per dettaglio.

raggiungere alla lettera. La stima individuale immediata del valore per il cliente è un limite, che richiederebbe una disponibilità illimitata in tempo reale di dati, risorse di calcolo e algoritmi, che soggiace a vincoli normativi (p.e. la difesa della privacy). Inoltre, la illimitata e istantanea capacità di variare i prezzi presenta anche problemi collaterali seri, quali la possibilità di *arbitraggio*[70] e di *cannibalizzazione*[71]. Queste considerazioni non intendono affatto contestare l'interesse del concetto di prezzo personalizzato, ma piuttosto inquadrarlo nei suoi oggettivi limiti. Peraltro, nella economia online continuano a verificarsi progressi verso questo limite ideale del prezzo personalizzato. Quello che sembra emergere è una tendenza verso una frammentazione spinta dei mercati, quindi una micro-segmentazione, che tende al limite verso lo *one-to-one pricing*, da intendere come punto asintotico della segmentazione.

Risulta evidente che per gli scopi del presente lavoro l'orientamento alla domanda nelle decisioni di prezzo è un presupposto imprescindibile, pur con i distinguo presentati e con il riconoscimento di un ruolo complementare dell'orientamento ai costi e alla competizione. Possiamo dire che l'orientamento alla domanda è parte del quadro concettuale in cui abbiamo definito l'ottimizzazione del reddito.

6.5 Domanda individuale e aggregata

Nel nostro modello la domanda è relativa a una certa combinazione di un prodotto, un mercato e un periodo di tempo. Chiamiamo per semplicità PMT questa combinazione. Modelliamo la domanda

---

[70] Si ha arbitraggio se un agente esterno può acquistare da noi il prodotto a un prezzo e rivenderlo a un nostro potenziale cliente a un prezzo maggiore, privandoci di una vendita addizionale (oltre ai possibili danni di immagine). Il tradizionale bagarinaggio dei biglietti per uno spettacolo è un esempio.

[71] Si ha cannibalizzazione se una vendita a profitto minore ci inibisce una vendita a profitto maggiore, dello stesso prodotto o di altri nostri. È come se inducessimo i nostri clienti a fare arbitraggio contro di noi.

come *domanda aggregata* di tutti i compratori che risulta dalle somma della *domanda individuale* di ciascun compratore. Il processo decisionale del compratore segue questo modello:

- In ogni periodo il compratore ha un prezzo di riserva (*reservation price*), che il venditore non conosce (anche se può formulare ipotesi al riguardo).
- In ogni periodo il compratore ha una e una sola opportunità di acquisto.
- In ogni periodo al compratore viene offerto un determinato prezzo (che può variare fra compratori e fra periodi)
- La decisone di acquisto del compratore consiste nell'accettare il prezzo offerto e comprare una e solo una unità di prodotto oppure rifiutarlo e non comprare alcuna unità di prodotto.
- Il compratore accetta di comprare se il prezzo offerto è minore o uguale al prezzo di riserva, rifiuta se è maggiore.

In una certa combinazione PMT, la domanda individuale del compratore i è dunque o 0 o 1, e vale 1 con una certa probabilità P_i. La domanda aggregata è il numero di compratori che accetta il prezzo offerto, quindi è la media delle probabilità individuali moltiplicata per il numero di compratori. Se la probabilità di acquisto è la stessa per tutti gli N compratori e vale P, allora la domanda aggregata è PN.

Questo modello ha un certo realismo solo per l'acquisto sporadico di beni durevoli e di prezzo elevato. Fortunatamente, si rivela efficace per molte applicazioni di Dynamic Pricing.

La domanda individuale è quindi una funzione del prezzo offerto dal venditore e anche del prezzo di riserva, che è ignoto al venditore. Il venditore può cercare di stimare il prezzo di riserva in base

- alle caratteristiche intrinseche del compratore, p.e. età, genere, residenza;

- al suo comportamento, p.e. la storia dei suoi acquisti, e anche dei suoi rifiuti di acqusito.

Siamo qui nel campo della profilazione del cliente e della predizione del comportamento individuale, dove possono giocare un ruolo importante le tecniche di predizione e classificazione viste altrove in questo lavoro.

Anziché cercare di predire la domanda individuale in funzione del prezzo, chiedendosi "*a questo prezzo, quanto è probabile che questo compratore accetti?*", il venditore può cercare di predire la domanda aggregata in funzione del prezzo, chiedendosi "*a questo prezzo, quanti compratori accetteranno?*". Sono questi due modi complementari di affrontare lo stesso problema di predizione della domanda.

6.6 Esempi di funzione di domanda

Descriviamo alcuni esempi di funzione prezzo-domanda che sono molto usati nella letteratura e trovano frequente applicazione anche nella prassi come ragionevoli ipotesi.

Indichiamo con *w(p)* la frazione di compratori che hanno prezzo di riserva uguale a p[72]. Ipotizziamo che *w* sia così distribuita tra i compratori:

$$w(p) = 10\% \ se \ 0 \leq p \leq 10$$

$$w(p) = 0 \ altrimenti$$

[72] La lettera *w* sta per *maximum willingness to pay*, massima disponibilità a pagare.

Se abbiamo un mercato potenziale di 20.000 compratori, ci sono 2.000 compratori disponibili a comprare a prezzo di 1 euro, altri 2.000 a prezzo di 1 o 2 euro, e così via fino a 2.000 compratori che sono disposti a pagare fino a 10 euro.

Se il venditore offre un prezzo *p* fra 0 e 10 euro, i compratori che accettano sono

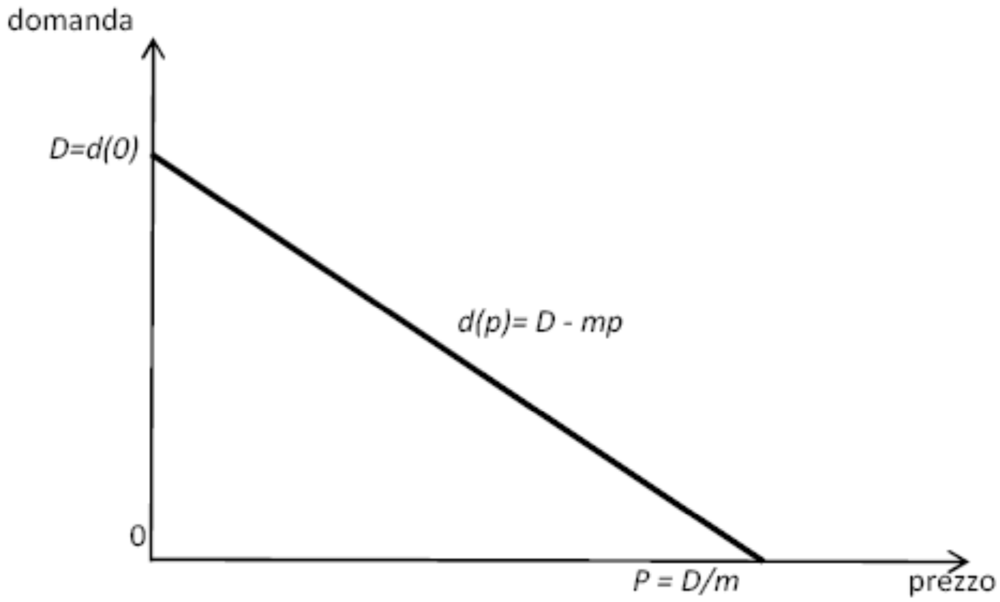
$$d(p) = 20.000 \sum_{x=p}^{10} w(x) = 20.000 \left(1 - \frac{p}{10}\right) = 20.000 - 2.000p$$

L'esempio illustra una regola generale: una domanda individuale distribuita uniformemente è associata a una domanda aggregata distribuita linearmente, e viceversa. Si capisce il perché notando che ad ogni incremento unitario di prezzo il numero di compratori che "abbandona il gioco", cioè che avrebbe comprato prima dello scatto ma non compra dopo, è una frazione costante (1/10 nell'esempio). Perciò la variazione di domanda aggregata è proporzionale alla variazione di prezzo, il che è per definizione un comportamento lineare.

La formula generale di una funzione di domanda aggregata lineare è

$$d(p) = D - mp$$

Quando il prezzo è zero la domanda raggiunge il livello massimo *D*, la dimensione dell'intero mercato (20.000 nell'esempio). Ad ogni incremento unitario di prezzo la domanda aggregata perde *m* compratori (2.000 nell'esempio). Quando *p* raggiunge il *prezzo di saturazione D/m* la domanda arriva a zero.



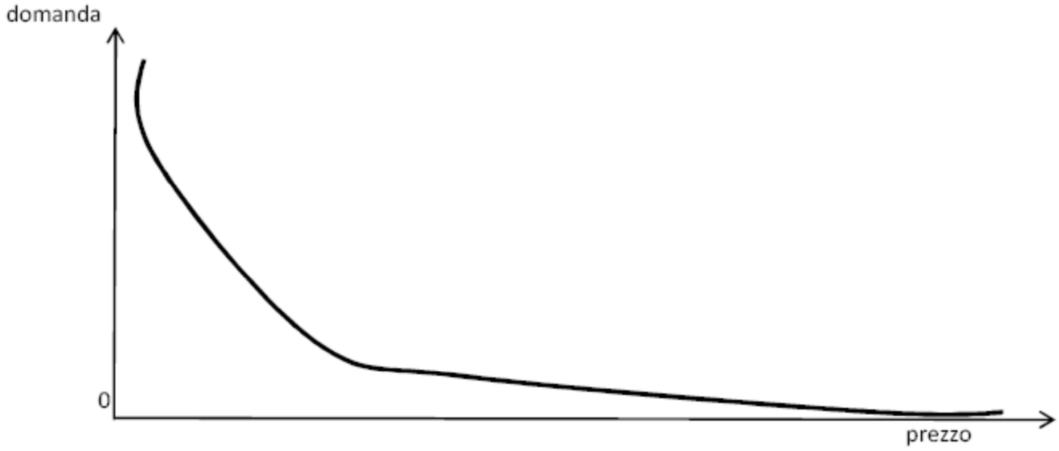
La funzione di domanda aggregata lineare non appare molto realistica. Ad esempio, se c'è un competitore allora è verosimile che quando il nostro prezzo si avvicina a quello del competitore l'effetto di una variazione di prezzo sia più forte di quando i due prezzi sono lontani. In questo senso la forma lineare appare più realistica come modello locale, per certi valori di prezzo, piuttosto che come modello dell'intera curva di prezzo-domanda.

La funzione di domanda aggregata a elasticità costante assume che l'elasticità

sia la stessa per ogni prezzo *p*. Anche questa funzione appare poco verosimile come modello globale della curva prezzo-domanda, mentre può rendere utilmente conto del comportamento della domanda in risposta a certe fasce di prezzo.

La funzione di domanda aggregata è

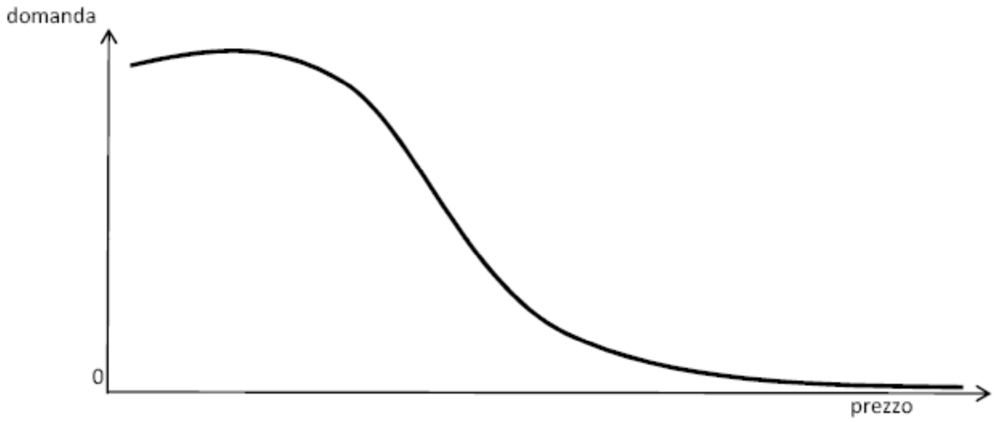
dove C è la domanda in corrispondenza del prezzo unitario.



La corrispondente funzione di domanda individuale è

Se l'impresa ha un competitore, o anche più competitori cha praticano pressi simili, allora deve confrontarsi con un prezzo di riferimento del mercato. Una curva di domanda realistica dovrebbe avere bassa elasticità a prezzi molto sopra o molto sotto il prezzo dei nostri competitori, e alta elasticità in vicinanza di quel prezzo. Il motivo è che un cliente molto o molto poco interessato al nostro prodotto non cambierà verosimilmente il suo atteggiamento in risposta a moderate variazioni di prezzo da parte nostra, mentre lo farà quando il nostro prezzo è vicino al prezzo di riferimento. Questa proprietà renderebbe la funzione di domanda più verosimile come modello globale, valido per tutta la curva prezzo-domanda. La più nota funzione di domanda aggregata con questa proprietà è la funzione logit:

I parametri *a* e *b* esprimono in forma lineare la sensibilità al prezzo, che raggiunge il massimo al prezzo −, che rappresenta il prezzo di riferimento del mercato di cui parlavamo prima.



La corrispondente distribuzione della willingness-to-pay, che rappresenta la domanda individuale, ha una forma a campana

Questa funzione di domanda è più realistica di quelle lineare e a elasticità costante quando si devono prendere in considerazione ampie variazioni del prezzo.

6.7 Customer commitment e costi incrementali

Un concetto importante per la comprensione del Dynamic Pricing è il *customer commitment* (impegno verso il cliente). Si ha un customer commitment quando il venditore si impegna con un cliente a vendergli un certo prodotto a un certo prezzo in un certo periodo. L'impegno può essere preso con un cliente singolo, con molti o con tutti i clienti di un intero mercato.

La forma più familiare di customer commitment è l'esposizione di un prezzo di listino. L'impegno è semplice nel caso di prezzo non negoziabile (tipico della distribuzione al dettaglio), più complesso se è prevista la possibilità di sconti (la regola negli impegni business-to-business). Ancora più complesso è il casso delle aste, in cui si concordano le regole di formazione futura del prezzo, ma non ancora il prezzo.

Il Dynamic Pricing, e anche il Capacity Control di cui parliamo altrove, possono essere pensati come metodi razionali per prendere sequenze di customer commitment con l'obiettivo di massimizzare il reddito.

Può sembrare che la valutazione dei costi sia un aspetto marginale del Dynamic Pricing, che pone problemi incomparabilmente più semplici della stima della curva di domanda o della conseguente individuazione dei prezzi ottimi. In realtà, la valutazione dei costi può essere un compito estremamente complesso e ricco di incertezze e componenti di giudizio soggettivo. Non a caso esistono così tanti metodi di valutazione dei costi e l'argomento resta al centro del dibattito nelle discipline economico-aziendali.

Nel Dynamic Pricing si utilizza un particolare concetto di costo il *costo incrementale*, definito come la differenza tra i costi totali che l'impresa ha se prende un customer commitment e i costi totali che ha se non lo prende. Si tratta quindi dei costi addizionali attribuibili al commitment. Il concetto ricorda il principio dello Activity Based Costing, in cui i costi vengono attributi alle attività dell'impresa. Ci ricolleghiamo così a quanto detto discutendo della equazione di equilibri economico: dobbiamo in qualche modo collegare costi e ricavi con le operazioni della gestione.

Una compagnia aerea che vende un volo a un cliente assume l'impegno a sopportare un costo incrementale che riunisce i costi di carburante, ristoro, commissioni da pagare per il servizio di prenotazione. Un negoziante di abbigliamento di moda che compra uno stock di capi a inizio stagione e non ha modo di restituirli o rivenderli su altro canale[73] avrà un costo incrementale pari a zero per la vendita di ciascun capo. In effetti, ormai il costo dello stock è un *costo inevitabile*. Se il negoziante vorrà praticare una politica di Dynamic Pricing, ad esempio una classica politica di sconto progressivo al progredire della stagione, potrà evitare di inserire i costi inevitabili nel suo processo decisionale. È per questo che nello studio del Dynamic Pricing i costi sono spesso visti come un problema secondario o talvolta assente: i costi inevitabili esistono e sono importanti, ma non sono rilevanti per prendere certe decisioni. Se capita che il prezzo ottimo scenda sotto il costo di acquisto, non è un buon argomento "No, facciamo un prezzo più alto perché questo non copre il costo". In realtà, coprire il costo non è un obiettivo perché il costo è già stato sostenuto in modo irreversibile: ora dobbiamo pensare solo al ricavo. A fini decisionali il coso può essere ignorato[74].

Il costo incrementale ha queste peculiarità:

- *È orientato al futuro*. Si valutano i costi futuri associati al commitment, non quelli passati e non reversibili. I costi inevitabili (*sunk costs*) non rientrano nei costi incrementali. Nel concetto di costi inevitabili rientrano non solo quelli passati e non reversibili, ma anche quelli futuri ma imputabili a decisioni passate e non reversibili.

[73] Per esempio nel canale outlet.

[74] Ciò non significa certo che possa essere ignorato a fini informativi: i costi inevitabili vanno in bilancio come gli altri!

- *È marginale*. Il coso incrementale è relativo a questo customer commitment, non alla somma o alla media di tutti i customer commitment.
- *Non è costo pieno*. I costi fissi di struttura non rientrano nel costo incrementale, perché non sono generati come costi addizionali di un commitment.
- *È incerto*. Essendo orientato al futuro, e dipendendo spesso dall'esito del commitment che lo genera, manifesterà il suo valore solo infuturo.

La differenza tra il prezzo di vendita di un prodotto e il suo costo incrementale è chiamata *margine*. La somma di tutti i margini di un prodotto, o mercato o periodo, o combinazione PMT, è chiamata *margine di contribuzione totale*. In molti casi l'equazione di equilibrio economico può essere utilmente riscritta (sempre a fini decisionali per l'ottimizzazione) come equazione della contribuzione totale

$$m(p) = (p - c)d(p)$$

cioè

$$Contribuzione\ totale = (Prezzo - Costo\ incrementale) \times Volume\ di\ vendita$$

L'obiettivo del decisore è qui espresso come massimizzazione della contribuzione totale.

Il problema di ottimizzazione non vincolata della contribuzione totale richiede di cercare il prezzo p^* che annulla la derivata della contribuzione totale. Semplici calcoli portano a questa condizione di ottimalità:

$$p^* d'(p^*) + d(p^*) = cd'(p^*)$$

Il membro sinistro dell'equazione non è altro che il ricavo marginale, mentre il membro destro è il costo marginale. Si vede ancora una volta all'opera il criterio marginalistico dell'ottimizzazione locale. Il *test di marginalità* richiede di verificare che a un certo livello di prezzo una piccola variazione al rialzo o al ribasso provoca un peggioramento della grandezza da ottimizzare, sia essa il reddito, il margine di contribuzione totale, il ricavo o qualsiasi altra.

Il test di marginalità evidenzia qualcosa di non immediatamente intuitivo se lo si applica alla ottimizzazione del ricavo. La grandezza da ottimizzare è qui

$$Ricavo = Prezzo \times Domanda$$

$$R(p) = pd(p)$$

Uguagliando a zero la derivata nel prezzo ottimale p^* si ottiene

$$d'(p^*)p^* + d(p^*) = 0$$

da cui

$$-\frac{d'(p^*)p^*}{d(p^*)} = \in(p^*) = 1$$

e quindi il prezzo che ottimizza il ricavo rende anche l'elasticità uguale a 1.

Abbiamo ottenuto due risultati distinti: il prezzo che massimizza il margine di contribuzione totale rende il ricavo marginale uguale al costo marginale, che è negativo, mentre il prezzo che massimizza il ricavo rende il ricavo marginale uguale a zero, quindi minore. Se il ricavo marginale è funzione decrescente del prezzo (e tipicamente lo è, almeno in prossimità del prezzo ottimale), allora *il prezzo che massimizza il ricavo è minore del prezzo che massimizza il margine di contribuzione*[75]. Occorre molta attenzione: massimizzare il ricavo non garantisce necessariamente che il margine di contribuzione sia elevato, anzi neanche che sia positivo, se il costo incrementale è maggiore di zero. La strategia di massimizzazione del ricavo può quindi presentare rischi inattesi.

[75] Nel caso estremo in cui il costo incrementale è zero, allora i due prezzi coincidono.

Un risultato interessante a cui portano queste considerazioni è che, in assenza di vincoli, il decisore è interessato ai prezzi maggiori o uguali al prezzo che massimizza il ricavo e minori o uguali al prezzo che massimizza il margine di contribuzione.

6.8 Surplus del compratore

Nella microeconomia svolgono un ruolo importate i concetti di *surplus del venditore* e *surplus del compratore*[76], che vengono studiati in relazione all'equilibrio del mercato, alla sua efficienza e al benessere generale. Qui diamo una interpretazione molto più semplice e limitata di questi concetti, destinata a chiarire il compito del decisore aziendale intento a ottimizzare il reddito.

Nel nostro modello ad ogni opportunità di acquisto un compratore compara il prezzo offerto dal venditore e accetta solo se questo è minore del suo prezzo di riserva. Se compra, il prezzo offerto diventa prezzo della transazione. La differenza fra il suo prezzo di riserva e il prezzo della transazione è il surplus del compratore. L'interpretazione è molto intuitiva: è il valore che il compratore "sente di avere risparmiato".

Analogamente, il venditore ha un suo prezzo di riserva, che è il prezzo minimo a cui è disposto a vendere, in base alle considerazioni fin qui svolte, che possiamo riassumere nel costo incrementale del commitment. La differenza fra il prezzo della transazione e il prezzo di riserva del venditore è allora il surplus del venditore, di nuovo con l'interpretazione intuitiva di importo che il venditore "ha guadagnato extra".

Poiché il prezzo di transazione è necessariamente intermedio fra i due prezzi di riserva, ne consegue che

---

[76] Citare Kreps p 315-seg.

$$Prezzo\ di\ riserva\ del\ compratore - Prezzo\ di\ riserva\ del\ venditore$$
$$= Surplus\ del\ compratore + Surplus\ del\ venditore$$

Possiamo ora vedere il Dynamic Pricing come l'uso della leva prezzo per massimizzare il surplus totale del venditore influenzando sia il surplus totale sia la quota di esso che va a surplus del venditore.

Si noti che questo enunciato non implica necessariamente la visione della transazione come gioco a somma zero. La differenza fra i prezzi di riserva[77] costituisce per ogni transazione un unico surplus da suddividere fra compratore e venditore. Si può allora pensare che il venditore voglia solo "avere la fetta di torta più grossa possibile" a danno del compratore. Ma questa è solo metà della storia: il Dynamic Pricing può anche aumentare il numero e il volume delle transazioni, aumentando il surplus totale e anche il surplus totale dei compratori. Non è necessario, ma è possibile, che il venditore "faccia crescere la torta" e quindi anche le fette che restano ai compratori.

Per comprendere meglio come il Dynamic Pricing possa aumentare il *welfare dei compratori* e quindi essere anche socialmente desiderabile, ricorriamo a un esempio[78].

Charlie è un consulente e due potenziali clienti, Alice e Bob, vorrebbero commissionargli uno stesso studio, che serve a entrambi. Alice è disposta a pagare fino a 700 euro per lo studio, Bob fino a 1.000 euro (questi sono i loro prezzi di riserva). Charlie ha un prezzo di riserva di 1.500 euro, al di sotto del quale preferisce non fare lo studio ma prendersi del tempo libero. Se Charlie deve richiedere lo stesso prezzo ai due potenziali clienti, lo studio non viene commissionato. Un prezzo fino a 700 euro procura a Charlie due commesse, ma un ricavo di al più 1.400 euro, insufficiente. Un prezzo tra 700 e 1.000 euro porta a Charlie solo la commessa di Bob, anch'essa insufficiente.

[77] Che si può suggestivamente interpretare come area della negoziazione.

[78] Tratto da Odlyzko.

Oltre i 1.000 euro non c'è alcuna commessa. Se invece Charlie può e vuole praticare due prezzi diversi, allora la soluzione sta nel chiedere 650 euro ad Alice e 950 euro a Bob. Charlie ha un ricavo di 1.600 euro e un surplus del venditore di 100 euro, Alice e Bob hanno ciascuno un surplus del compratore di 50 euro e tutti restano più contenti che nell'ipotesi di prezzi uniformi.

6.9 Discriminazione di prezzo

La *discriminazione di prezzo* consiste nella offerta di un prezzo diverso per diverse occasioni di acquisto di uno stesso prodotto. Clienti diversi nello stesso momento riceveranno proposte di prezzo diverse, e così lo stesso cliente in momenti diversi.

Si usa tradizionalmente la classificazione delle forme di discriminazione di prezzo in tre *gradi*, seguendo Pigou[79], anche se il fenomeno è estremamente articolato e sfugge a classificazioni molto precise.

Nella *discriminazione di prezzo di terzo grado* a gruppi diversi di clienti si offrono prezzi diversi. Esempi ben noti sono trasporti e spettacoli offerti a prezzi più favorevoli ad anziani, giovani, donne o altre categorie. Al di là delle motivazioni etiche o normative, praticare prezzi diversi può anche avere motivazioni di massimizzazione del reddito. Vendendo biglietti scontati a giovani e anziani, che in media hanno maggiore elasticità al prezzo, si possono riempire posti al cinema che altrimenti resterebbero vuoti, in definitiva aumentando il reddito.

Le caratteristiche su cui si basa la separazione in gruppi possono essere intrinseche ai clienti, come età o genere, ma possono anche essere dei comportamenti spontanei o sollecitati dal venditore stesso. Per esempio, i coupon sono effettivamente utilizzati da clienti ad alta elasticità della domanda (che rispondono bene agli sconti), e quindi il loro uso è un rivelatore dell'appartenenza a

[79] Citazione, e anche Philips su Pigou.

gruppi. Per questo i coupon funzionano anche quando non sono riservati a gruppi ma sono accessibili a chiunque, perché basta ritagliarli da un giornale o scaricarli da Internet e stamparli. La cosa davvero importante, in ultima analisi, non è l'appartenenza a un gruppo: fintanto che chi sfrutta gli sconti ha domanda più elastica di chi non li sfrutta, la discriminazione di prezzo funziona. Detto in un altro modo: stiamo dividendo i clienti in due gruppi facilmente distinguibili, chi usa e chi non usa i coupon, e questi gruppi hanno effettivamente funzioni di domanda diverse, che danno opportunità di ottimizzazione.

Tali pratiche sono diffuse, ma non quanto ci si potrebbe attendere a prima vista. In realtà, ci sono potenti forze che possono rendere difficile o controproducente strategie di questo tipo. Possiamo sintetizzarle in tre tipi di ostacoli:

1. *Segmentazione imperfetta*. Il gruppo alla base della discriminazione di prezzo potrebbe non avere il comportamento atteso da chi progetta la discriminazione stessa. Potremmo quindi offrire prezzi minori a chi in realtà ha una *willingness to pay* più elevata.
2. *Cannibalizzazione*. La discriminazione di prezzo incentiva chi ha una elevata *willingness to pay* a dissimularla e tentare di accedere ai prezzi riservati ai gruppi che ne hanno una più bassa.
3. *Arbitraggio*. Soggetti terzi sono incentivati a cercare modi per acquistare il prodotto a prezzo basso per rivenderlo a clienti ad alta *willingness to pay* a prezzo più elevato ma comunque minore di quello riservato al loro gruppo, trattenendo per sé il differenzial edi prezzo.

In sintesi, è difficile operare una discriminazione corretta, ed anche se si riesce c'è il rischio che sia aggirata. Resta vero che se si riesce a evitare questi tre problemi, la discriminazione di terzo grado, basata sulla suddivisione in gruppi può essere un potente strumento di ottimizzazione del reddito.

Nella *discriminazione di prezzo di secondo grado* ad ogni cliente si offre una gamma di proposte tra le quali può liberamente scegliere. Esempi sono le classi *first*, *business* e *economy* dell'aereo per una compagnia tradizionale (non *low cost*), oppure le versioni *basic*, *professional* e *enterprise* di un prodotto software. Diffusissimi sono gli schemi di prezzo *non lineari*, nei quali il prezzo dipende dalla quantità: sconti al crescere dei volumi, viceversa sconti per le prime unità di prodotto, quota fissa di ingresso e poi prezzi standard, e infiniti altri. L'estrema complessità delle tariffe praticate dalle compagnie telefoniche è esperienza di tutti.

In questo tipo di discriminazione il cliente si *auto-seleziona* come appartenente a un gruppo. Anche se si usa comunemente la distinzione tra *group pricing* e *product versioning*, il confine tra i due schemi di prezzo non è ben definito. L'idea comune è di stimare la *willingness to pay* del cliente e collocarlo in un gruppo al quale proporre un certo prezzo. Nel group pricing la classificazione in un gruppo avviene con l'osservazione delle caratteristiche del cliente. Nel product versioning non ci si limita all'osservazione, ma si avanzano delle proposte che provocano dei comportamenti di auto-selezione. Lo strumento dei coupon può essere visto come discriminazione di terzo grado se si pensa al cliente che ritagliando il giornale manifesta la sua appartenenza a un gruppo sensibile al prezzo, perché accetta il disagio di ritagliare il coupon, conservarlo e presentarlo. Ma può anche essere visto come discriminazione di secondo grado se si immagina che il cliente scelga fra il prodotto profumo a prezzo 100 e il prodotto profumo+disagio a prezzo 90.

Nella *discriminazione di prezzo di primo grado* ad ogni cliente si offre un prezzo personalizzato, valido solo per lui e solo in quel momento (si usano spesso le dizioni *one-to-one pricing* o *customer driven pricing*). Idealmente il prezzo dovrebbe approssimare la *willingness-to-pay* del cliente, per massimizzare il surplus del venditore a scapito del surplus del compratore. Questo schema, che appare come un Santo Graal del pricing, può essere approssimato in certa misura nel commercio di massa grazie alla diffusione di Internet e dei metodi di data mining e ottimizzazione.

6.10 Tipologie di decisioni del Revenue Management

Da questa sezione passiamo a considerare il contesto economico del Revenue Management, inteso come comprensivo del Dynamic Pricing e del Capacity Control. Le considerazioni che svolgeremo si riferiscono in primo luogo al Revenue Management nella sua concreta veste “storica”, ma si possono estendere, anche se non sempre e non esattamente, anche all’ottimizzazione del reddito in un senso più generale.

Nel Revenue Management si prendono tre tipi di decisioni:

1. *Decisioni strutturali*. Si decide l’assortimento di prodotti in vendita, compresi i *bundle*, i pacchetti di prodotti; il formato di vendita, p.e. listino o aste; la strategia di segmentazione del mercato; le condizioni contrattuali di vendita, p.e. i rimborsi.
2. *Decisioni di prezzo*. Si decide la struttura dei prezzi e la sua dinamica nel tempo, p.e. la politica di sconti.
3. *Decisioni di quantità*. Si decide la quantità di prodotti disponibile per la vendita in un certo periodo per un certo segmento di mercato; la politica di accettazione o rifiuto di un ordine.

In genere un’azienda dovrà prendere decisioni di tutti i tre tipi, ma quale tipologia risulti in concreto la più importante dipende dal settore e dalle scelte strategiche dell’azienda. Per le compagnie aeree low cost le decisioni di prezzo sono essenziali mentre quelle di capacità tendono a svanire, perché non esistono le classi di posti. Viceversa, nelle compagnie tradizionali, la struttura dei prezzi è meno articolata e molto meno dinamica, ma in compenso assume un ruolo importante la distinzione in classi di posti, e quindi la gestione della leva capacità.

6.11 Eterogeneità e incertezza della domanda

Non è solo un caso se storicamente il Revenue Management si è sviluppato molto più in alcuni settori industriali che in altri. In effetti, la sua efficacia dipende dal grado in cui sono soddisfatti alcuni requisiti.

Il primo requisito è la *eterogeneità della domanda*. Più i compratori mostrano tale eterogeneità, più è possibile suddividerli in segmenti fortemente differenziati e quindi sfruttare tali differenze. L'azienda potrà meglio variare i prezzi fra segmenti per avvicinarli alla willingness-to-pay di ciascuno e quindi estrarre in maggiore misura il surplus del compratore. Questa possibilità non è certo automatica e scontata, dovendo l'azienda contrastare la cannibalizzazione e l'arbitraggio ed evitare conseguenze legali o di impopolarità della differenziazione di prezzo, ma laddove sia possibile si esaltano le potenzialità dell'ottimizzazione.

Il concetto di eterogeneità della domanda non si esaurisce nel concetto di possibilità di segmentazione inteso in senso statico, ma lo estende alla dinamica temporale: una domanda eterogenea può essere tale perché variante fra individui nello stesso tempo, così come in tempi diversi per gli stessi individui.

Analogamente, la domanda è eterogenea se il comportamento di acquisto dei compratori varia in dipendenza dal canale di vendita. Per esempio, un albergo potrà decidere politiche di prezzo e quantità per i due canali offline e online. Oppure potrà assegnare lotti di camere a una agenzia e gestire in proprio le altre.

Ancora, prezzo e quantità potranno essere variati in dipendenza dalla dimensione spaziale: non solo prezzi e quantità differenziate per mercati locali, in senso tradizionale, ma addirittura per posizione istantanea del compratore, se si pensa ad acquisti online in mobilità.

In generale, l'ottimizzazione sarà tanto più applicabile con efficacia quanto più sarà possibile identificare delle variabili di qualunque genere che introdotte in un modello permettano di prendere decisioni differenziate su prezzo e quantità. È questo un concetto molto astratto, che però una volta istanziato su scelte concrete è coerente con il classico concetto di segmentazione e apre entusiasmanti possibilità, che già oggi molte aziende stanno esplorando e in alcuni casi sfruttando con grandissimo profitto. Nel capitolo relativo al nostro modello decisionale, il tema dell'eterogeneità della domanda è trattato a un livello astratto e generale.

Un requisito analogo all'eterogeneità è l'*incertezza della domanda*. Tanto più la domanda è incerta, difficile da prevedere, di natura prettamente stocastica, variabile nel tempo, tanto più una gestione scientifica di prezzo e capacità può apportare benefici all'azienda rispetto a una gestione empirica.

A ben vedere, la stessa eterogeneità della domanda può essere vista in termini di incertezza: se le domande dei diversi contesti sono molto diverse, allora la distribuzione delle distribuzioni stocastiche delle risposte è a bassa incertezza, vale a dire ha un forte valore informativo.

L'industria del trasporto aereo mostra una elevata incertezza della domanda, in specie per l'alta variabilità nel tempo, ma anche rispetto ai segmenti di viaggiatori. Questa incertezza può essere affrontata manovrando l'una o l'altra delle leve di quantità e prezzo.

6.12 Vincoli di produzione

Se il venditore può facilmente assorbire la variabilità e l'incertezza della domanda con un facile, rapido ed economico ripristino della capacità di vendita, allora il problema dell'ottimizzazione risulta facilitato, se non altro rispetto al controllo della quantità: al manifestarsi della domanda si risponde di volta in volta producendo ciò che serve nella quantità che serve. Stiamo parlando di produzione in senso lato: la capacità può essere ripristinata anche con l'approvvigionamento da altri soggetti.

La produzione può però essere sottoposta a vincoli rigidi: tempi di produzione, costi fissi, economie di scala, costi di commutazione tra produzioni diverse, vincoli di magazzino, difficoltà di trasporto, oneri finanziari, scarsa liquidità, possono rendere molto complesso il ripristino della capacità di vendita. In generale, la rigidità di produzione costringe a frequenti decisioni complesse rispetto a contesti diversi, dovendo distribuire le risorse fra i periodi di tempo, i segmenti di clientela, i canali eccetera. Tanto meno flessibile è la produzione, per vincoli tecnico-fisici, logistici o economico-finanziari, tanto più una ottimizzazione condotta con criteri scientifici diventa preziosa. In condizioni di capacità scarsa e non ripristinabile facilmente, è importante riuscire a spostare opportunamente la domanda su periodi e segmenti, usando il prezzo o la capacità.

Nel trasporto aereo la capacità è molto rigida, poiché il numero di posti su un velivolo non può essere cambiato, non nel breve periodo. Meglio, può essere cambiato solo con una complessa organizzazione che consenta di scegliere il velivolo per una tratta proprio in base agli andamenti delle prenotazioni e vendite di biglietti. Tale soluzione è possibile, ma a sua volta costosa. Se la capacità massima è prefissata, allora è evidente l'importanza di dirigere la domanda con manovre di prezzo e quantità.

Non solo la capacità è fissa nel breve periodo, ma non esiste magazzino, non potendo conservare i posti invenduti su un volo per un volo successivo. In definitiva, quando una compagnia aerea si impegna a effettuare certi voli su una tratta, essa prefissa il suo output di produzione per lungo tempo e in ogni distinto istante, senza flessibilità di capacità e senza magazzino. I costi sono quindi prefissati con largo anticipo, molto prima che la domanda si manifesti, e in definitiva sono pressoché interamente costi fissi. Questo rende estremamente importante lo sfruttamento di ogni unità di bene, cioè di ogni posto vendibile, poiché il costo unitario del venduto dipende in modo fortissimo dal volume di vendita.

6.13 Altri requisiti

Il vantaggio di modelli scientifici di ottimizzazione è maggiore quando il prodotto è deperibile, nel senso che perde valore con il tempo. Nel trasporto aereo il valore di un posto su un volo si azzera al momento della partenza, nell'industria alberghiera il valore di una camera scende rapidamente verso zero in prossimità della sera, così come un posto a teatro o allo stadio. Anche il valore di merci alimentari ha un calo con il tempo, anche se in modo meno marcato, e ancora con un altro ritmo cala nel tempo il valore dell'abbigliamento di moda stagionale.

La *deperibilità del prodotto* fa sì che la gestione della capacità cresca di importanza, e in tal caso anche la leva di prezzo può diventare importante.

Anche il meccanismo delle *prenotazioni* è un fattore che aumenta l'importanza dell'ottimizzazione mediante il prezzo e la quantità. Il tempo di prenotazione è importante come segnale dell’andamento della domanda, come indice delle preferenze temporali dei compratori, come criterio di segmentazione.

Nel mercato dei beni di lusso il prezzo ha una funzione di “segnale di qualità”. Per esempio il prezzo elevato di una borsa di gran marca è di per sé parte del valore, in quanto contribuisce al valore di status symbol. In un senso diverso, la parcella di un medico o di un avvocato suggeriscono il valore del professionista. In contesti di questo genere, in cui prezzo e qualità sono associati nella mente del compratore, la manovra sui prezzi può essere rischiosa fino al punto da essere inattuabile.

L’ottimizzazione mediante manovre sul prezzo è quindi molto più praticabile quando il prezzo non è percepito come indicatore di qualità.

Problemi di genere diverso possono nascere quando la discriminazione di prezzo è vista come “ingiusta” o anche illegale, come si è detto nel capitolo sul Dynamic Pricing. Anche in questi casi le aziende, comprensibilmente, esitano a introdurre manovre sui prezzi che pure potrebbero essere estremamente utili.

Nell’industria del trasporto aereo, specie dopo l’ascesa delle compagnie *low cost*, la percezione del prezzo non è molto legata alla qualità, e infatti le compagnie aeree utilizzano ampiamente manovre di prezzo.

Un fattore decisivo per l’adozione di metodi di ottimizzazione del reddito è la *cultura manageriale* dell’azienda. Infatti, è richiesta non soltanto l’apertura verso tecnologie informatiche non banali, ma anche la disponibilità a rivedere abitudini consolidate, e soprattutto la disponibilità a lasciare che sistemi automatici e modelli matematici supportino le decisioni, o che addirittura le prendano essi stessi, come nel caso dei sistemi online. Tutti requisiti non facili da soddisfare nella gran parte delle aziende.

Infine, ma non certo ultimo per importanza, occorre disporre di sistemi informatici adeguati, che permettano di raccogliere molti dati e di elaborarli molto velocemente. Questo requisito è però destinato a diventare meno esigente con il progredire delle tecnologie informatiche e con il crescente uso di Internet per transazioni di acquisto.

7 OTTIMIZZAZIONE NEL MODELLO DECISIONALE

In questo capitolo descriveremo alcuni metodi di ottimizzazione applicabili a problemi rappresentati nel nostro modello decisionale.

7.1 Condizioni di ottimalità

Abbiamo definito una politica come una funzione stocastica che assegna ad ogni azione una probabilità di essere scelta in un certo stato:

$$\pi(s, a) = \Pr(a_t = a | s_t = s)$$

Per definire i criteri di ottimalità ci è utile conoscere il *valore di uno stato*, che è il valore atteso scontato della serie di flussi generati dal processo decisionale sotto la politica π partendo dallo stato *s*. Definiamo il valore scontato della serie di flussi come

$$R_t \stackrel{\text{def}}{=} \sum_{k=0}^{T-t} \gamma^k r_{t+k}$$

e il valore dello stato come

$$V^\pi(s) = E_\pi\{R_t | s_t = s\}$$

L'operatore di valore atteso E_π dà la media dei possibili valori del suo argomento pesati con le rispettive probabilità assegnate dalla politica π.

Una definizione simile assegna un valore a una coppia stato-azione:

$$Q^\pi(s, a) = E_\pi\{R_t | s_t = s, a_t = a\}$$

Le due funzioni di valore possono essere stimate in base all'esperienza. Se seguiamo la politica π e memorizziamo la media dei valori ottenuti in seguito all'entrata in un certo stato, questa è una stima campionaria di $V^\pi(s)$; se manteniamo i dati disaggregati per stato e azione allora abbiamo una stima campionaria di $Q^\pi(s, a)$. Su grandi campioni queste stime convergono verso i valori "veri" delle due funzioni. Questa procedura è la base dei *metodi Monte Carlo*, che vedremo in seguito. È però chiaro che non sempre si ha la possibilità di raccogliere dati sufficienti per provare tutti gli stati o, ancora meno, tutte le combinazioni stato-azione. Un'alternativa è definire due funzioni parametriche e usare l’esperienza per adattare i parametri ai valori osservati. Questa è l’idea della *approssimazione di funzioni*, che a noi vediamo come *generalizzazione dell’informazione*.

Si dimostra che la funzione di valore *V* si può esprimere anche così:

$$V^{\pi}(s) = \sum_{a} \pi(s,a) \sum_{s'} \sigma(s,a,s')[\rho(s,a,s') + \gamma V^{\pi}(s')]$$

Per ogni stato possibile successore di *s* si prenda la probabilità di scegliere una azione e di entrare nello stato successore; queste probabilità congiunte servono da pesi nel calcolo del valore atteso, in cui si considerano insieme i flussi immediati dati da ρ e i flussi futuri scontati sintetizzati da *V*. Il tutto in base alla politica π fissata.

Questa importante formula è nota come *equazione di Bellman*. Si noti come questa equazione sia *ricorsiva*, nel senso che esprime la funzione mediante la funzione stessa, e proceda *all’indietro*, nel senso che il valore della funzione in uno stato è definito mediante il valore in uno stato successivo. La ricorsione all’indietro implica che per conoscere il presente dobbiamo già conoscere il futuro. Perciò, dobbiamo avere già una stima di valori futuri, quindi una *previsione* approssimata di quello che potrà accadere. L’equazione di Bellman è uno strumento matematico essenziale nella teoria dei processi decisionali markoviani. La sua soluzione può comportare una seria complessità computazionale e problemi di disponibilità di dati sufficienti. I metodi di ottimizzazione che vedremo sono tutti volti a risolvere questa equazione per approssimazioni successive.

Ottimizzare il processo decisionale significa trovare la *politica ottimale* π^*, che è migliore o uguale a ogni altra politica su ogni stato

$$V^{\pi^*}(s) \geq V^{\pi}(s) \forall s$$

Ci possono essere molte politiche ottimali, che hanno la stessa funzione di valore di stato

$$V^*(s) = \max_{\pi} V^{\pi}(s) \;\; \forall s$$

e la stessa funzione di valore di stato-azione

$$Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a) \;\; \forall s,a$$

Vale la relazione

$$Q^*(s,a) = E\{r_t + \gamma V^*(s_{t+1}) | s_t = s, a_t = a\}$$

Scriviamo quindi le condizioni di ottimalità di Bellman per V^* e Q^*:

$$V^*(s) = \max_a \sum_{s'} \sigma(s,a,s')[\rho(s,a,s') + \gamma V^*(s')]$$

$$Q^*(s,a) = \sum_{s'} \sigma(s,a,s')[\rho(s,a,s') + \gamma \max_{a'} Q^*(s',a')$$

Se $\gamma < 1$ resta garantito che esiste un valore che soddisfa queste equazioni.

Le condizioni non fanno riferimento a una specifica politica, ma presuppongono solo che essa sia ottimale. Il principio è che una politica ottimale è tale a ogni passo: ogni decisione massimizza la somma di un flusso immediato, imputabile a una singola azione, e di una serie di flussi futuri, imputabili alla politica ottimale stessa applicata a partire dalla prossima decisione. Una conseguenza di questo principio è che la decisone ottimale in un certo passo non dipende da tutti i possibili percorsi successivi, ma solo da quelli ottimali. In questo modo il numero di cammini da considerare diminuisce molto, con conseguente beneficio computazionale. Possiamo dire che un decisore razionale prende la decisione presente confidando di essere razionale anche in futuro.

7.2 Soluzione esplicita dell'equazione di Bellman

La condizione di ottimalità per V è in effetti un sistema di N equazioni lineari in N incognite, se N è il numero di stati. Se si conosce la dinamica del processo, cioè le funzioni σ e ρ, allora il sistema di equazioni può essere risolto con un grande varietà di tecniche. Lo stesso vale per Q^*.

Una volta che si conosce V^*, diventa relativamente semplice determinare una politica ottimale. Per ogni stato s ci saranno una o più azioni ottimali; ogni azione che assegna probabilità non nulle a tutte e sole queste azioni (non importa quali probabilità) è una politica ottimale. Questo significa che in un certo stato le azioni che appaiono migliori "guardando un solo passo avanti" sono in effetti le migliori. La funzione V^* "memorizza il futuro", sintetizzando una serie di flussi in un'unica grandezza, ed è questo che rende possibile l'ottimizzazione con orizzonte di pianificazione di un solo passo.

Se si conosce Q^* non c'è neanche bisogno dell'unico passo di pianificazione: in ogni stato s si sceglie semplicemente una delle azioni che massimizzano $Q^*(s, a)$, vale a dire che si pianificano zero passi. Il motivo è lo stesso: la funzione memorizza già le conseguenze di un singolo passo avanti. In entrambi i casi il futuro sembra sparire perché è sintetizzato in una grandezza presente, così come gli stati sintetizzano il passato. È all'opera la proprietà di Markov, per la quale non è necessaria una *ricerca globale* sugli stati, ma è sufficiente una *ricerca locale*.

Tuttavia, la soluzione esplicita dell'equazione di Bellman richiede ugualmente l'esame di tutte i percorsi possibili, perché V^* e Q^* sono a loro volta calcolate così. La risoluzione esplicita dell'equazione di Bellman è davvero utile se si verificano queste tre condizioni:

1. conosciamo accuratamente la dinamica del processo;
2. abbiamo sufficienti risorse computazionali;
3. vale la proprietà di Markov.

Spesso una o più di queste assunzioni non vale e la soluzione esplicita non è una soluzione utile e percorribile. Si può allora ricorrere a metodi di soluzione approssimata, come quelli che stiamo per vedere.

7.3 Programmazione Dinamica

Il termine *Programmazione Dinamica* (DP) denota una metodologia e un insieme di algoritmi per la soluzione di una ampia gamma di problemi, fra cui i processi decisionali markoviani. Le caratteristiche di questa metodologia sono già state accennate parlando delle equazioni di Bellman, il fondatore della DP:

- *Ricorsività*. Un problema è suddiviso in più problemi, alcuni risolvibili immediatamente, altri che sono simili al problema originario ma di ridotta complessità. Il procedimento si ripete finché tutti i problemi sono ridotti in modo immediato. A quel punto le soluzioni dei problemi meno complessi sono ricomposte per risolvere i problemi più complessi, e così a ritroso fino al problema originario.
- *Uso di stati*. Un numero ristretto di variabili di stato contiene tutta l'informazione necessaria per affrontare un problema.
- *Conoscenza della dinamica di sistema*. È noto in quale modo i problemi si generano gli uni dagli altri.

- *Bootstrapping*. Anche le soluzioni si generano le une dalle altre. Occorre sapere già qualcosa su come trovare soluzioni, per poter “propagare” le soluzioni stesse[80].
- *Semplificazione della ricerca*. Non si esplorano tutte le alternative, ma solo quelle che possono essere componenti di percorsi risolutivi ottimali.

Riprendiamo la definizione di valore di stato:

$$V^{\pi}(s) = \sum_{a} \pi(s,a) \sum_{s'} \sigma(s,a,s')[\rho(s,a,s') + \gamma V^{\pi}(s')]$$

Se conosciamo completamente la dinamica di sistema data dalle funzioni σ e ρ, possiamo certamente risolvere il sistema di equazioni lineari, compito che però può essere molto pesante in termini di risorse computazionali richieste. In alternativa, possiamo ricorrere alla approssimazione di funzioni.

Partiamo con una funzione V_0 definita in modo arbitrario, e poi generiamo una sequenza V_1, V_2, … di funzioni che approssimano iterativamente la funzione V^{π}. Ogni approssimazione è generata dalla precedente con questa regola di aggiornamento:

$$V_{k+1}(s) = E_{\pi}\{r_t + \gamma V_k(s_{t+1}) | s_t = s\} = \sum_{a} \pi(s,a) \sum_{s'} \sigma(s,a,s')[\rho(s,a,s') + \gamma V_k(s')]$$

Se $\gamma < 1$, allora la sequenza di approssimazioni converge verso il vero valore di V^{π}, che resta immutato quando sottoposto alla regola di aggiornamento. Per scopi pratici, si può sospendere il procedimento quando le iterazioni modificano di molto poco il valore V, accontentandosi di approssimarlo nel grado desiderato.

Ad ogni iterazione il valore di uno stato viene aggiornato con una ricerca sugli stati raggiungibili in un passo, oltre che con il flusso immediato. La ricerca globale (che consisterebbe nel guardare avanti fino alla fine dei possibili cammini) è sostituita da una iterazione di ricerche locali.

Con questo metodo possiamo valutare il valore di una politica. Il nostro obiettivo è di trovare una politica ottimale e quindi la valutazione di singole politiche è certamente utile, però valutare semplicemente tutte le possibili politiche e scegliere quella con il massimo valore di V^{π} non è in

---

[80] *Bootstrapping* (soolevarsi tirandosi su per le stringhe delle scarpe) è un termine per indicare che un processo prende l’avvio grazie a un processo simile più semplice.

generale un metodo perseguibile, perché il loro numero può essere enorme. Si ricorre allora a un metodo iterativo che partendo da una politica iniziale raggiunge una politica ottima con una serie di aggiustamenti successivi.

Per semplificare la descrizione, limitiamoci alle politiche *deterministiche*, cioè tali che in ogni stato una azione viene scelta con probabilità 1, tutte le altre con probabilità 0. Scegliamo in modo arbitrario una politica iniziale π_0.

Ora prendiamo un qualsiasi stato s e consideriamo l'azione a che π_0 sceglie sempre quando lo stato è s una azione diversa a' che non viene scelta mai:

$$\pi_0(s, a) = 1$$

$$\pi_0(s, a') = 0$$

Valutiamo le due azioni e verifichiamo se

$$Q^{\pi_0}(s, a') > V^{\pi}(s)$$

Se la relazione non vale cerchiamo un'altra azione oppure cambiamo stato. Se in nessun caso la relazione vale, allora la politica π_0 è già ottimale. Se invece la relazione vale, allora creiamo una nuova politica π_1 che differisce da π_0 soltanto per questi due assegnamenti

$$\pi_1(s, a) = 0$$

$$\pi_1(s, a') = 1$$

La nuova politica si differenzia dalla precedente perché nello stato s sceglie a' anziché a. Anche π_1 è deterministica e vale

$$V^{\pi_1}(s) > V^{\pi_0}(s)$$

perché la nuova politica si comporta meglio della precedente in uno stato e allo stesso modo in tutti gli altri.

Il procedimento è iterato fino a trovare una politica deterministica ottimale:

$$V^{\pi'}(s) > V^{\pi}(s) \forall s$$

Questo metodo può essere generalizzato a politiche non deterministiche, con alcuni tecnicismi che tralasciamo.

Il miglioramento iterativo della politica può essere molto dispendioso di risorse computazionali, ma esistono metodi per approssimarlo con un numero limitato di calcoli.

È convinzione abbastanza diffusa che la DP sia computazionalmente poco efficiente, ma non ci sono basi scientifiche per questa idea. Anzi, poiché il procedimento tende a escludere dalla ricerca di soluzioni i passi che non fanno parte di cammini ottimali, spesso la DP si rivela più efficiente di metodi meno rigorosi.

Il vero limite pratico della DP è piuttosto il requisito di conoscenza della dinamica del sistema. Se questo requisito è soddisfatto, il metodo della DP offre un quadro matematico attraente, con soluzioni ottimali e garanzie di convergenza, anche se in pratica il ricorso ad approssimazioni può offuscare la purezza matematica, e con buona efficienza. Ma in molte applicazioni il requisito non è soddisfatto, neanche in modo approssimato. Si ricorre allora a metodi euristici, che danno meno garanzie ma sono meno esigenti in termini di informazione necessaria. I metodi Monte Carlo rispondono a queste esigenze.

7.4 Metodi Monte Carlo

I *Metodi Monte Carlo* (MCM) sono una classe di metodi basati sul principio di operare non con una distribuzione stocastica integralmente nota ma con suoi campioni. Per i nostri problemi questo significa che non abbiamo bisogno di conoscere completamente la dinamica del sistema data dalle funzioni σ e ρ in forma di equazioni esplicite (non c'è *bootstrapping*), ma ci basta avere a disposizione delle sequenze di stati, azioni e flussi. Questi dati potranno essere reali o simulati. Utilizzare dati reali significa basarsi unicamente su quello che si apprende "sul campo" senza avere conoscenza pregressa. Per utilizzare dati simulati occorre invece disporre di un meccanismo per generare le sequenze. Questo sembra riportarci indietro al problema di conoscere già la dinamica del sistema, ma in realtà non è così. Spesso si riesce a generare dati di transizione fra stati e flussi anche se non si sa esprimerne le leggi in forma esplicita.

L'idea di base è semplice: se sappiamo con quale *frequenza* si sono verificati in passato certi eventi, ci aspettiamo che in futuro quegli eventi si presentino con analoghe *probabilità*.

Vediamo l'applicazione del metodo all'apprendimento del valore degli stati *V* per una certa politica π.

Facciamo partire il processo, sia esso reale o simulato, seguendo π. Registrando transizioni, di stato e flussi siamo in grado di dire quale è stata la somma dei flussi scontati a partire dal momento in cui il processo ha attraversato lo stato *s*. Questa somma è considerata un campione di *V(s)*. Al ripetersi di esecuzioni del processo decisionale, accumuleremo nuovi campioni. In ogni momento stimiamo *V(s)* con la media dei valori dei campioni raccolti fino al momento. Nel tempo questa stima convergerà verso il vero valore di *V*.

Più formalmente, se s_t è lo stato visitato al tempo *t* durante un episodio di esecuzione del processo decisionale, la nuova stima del suo valore $V(s_t)$ è data da

$$V(s_t) \leftarrow V(s_t) + \alpha[R_t - V(s_t)]$$

dove $\alpha < 1$ è un parametro. Questa regola di aggiornamento può essere così tradotta in parole

$$Nuova\ stima \leftarrow Vecchia\ stima + Passo[Obiettivo - Vecchia\ stima]$$

L'obiettivo è la grandezza che dovremmo aggiungere alla vecchia stima affinché la nuova coincida con i dati osservati; infatti R_t è proprio il valore che cui $V(s_t)$ deve stimare. Il passo è un regolatore della velocità di apprendimento: avviciniamo la nuova stima all'obiettivo un po' più della vecchia, ma non completamente, per evitare che a ogni iterazione la stima oscilli troppo e far sì che la stima si avvicini dolcemente alla congruenza con i dati empirici. Può sembrare innaturale, ma sarà più comprensibile quando discuteremo del problema dell'*overfitting*.

In modo analogo stimeremo *Q(s, a)* per ogni stato ed azione.

Vediamo uno schema di base per una concreta implementazione. Non è affatto l'unico schema possibile, ma ha il vantaggio di una certa semplicità intuitiva e mette in mostra i concetti di base utilizzati da altri metodi più sofisticati.

1. *Inizializzazione*. Fissiamo tutti i valori *Q(s, a)* in modo arbitrario. Fissiamo una politica π arbitraria. Per ogni coppia *(s, a)* creiamo una lista vuota di somma di flussi scontati.
2. *Episodio sperimentale*. Seguiamo π e registriamo la sequenza di flussi generati e stati raggiunti.
3. *Aggiornamento di Q*. Finito l'episodio, associamo ad ogni coppia *(s, a)* la somma dei flussi scontati a partire dalla prima occasione in cui si è eseguito *a* in *s*. Aggiungiamo questo valore alla lista esistente. Ricalcoliamo la media dei valori nella lista e assegniamo questo valore a *Q(s, a)*.
4. *Aggiornamento di π*. Modifichiamo i valori *π(s, a)* in funzione dei nuovi valori *Q(s, a)*.
5. *Ripetizione*. Ad ogni nuovo episodio di esecuzione del processo ripartire dal passo 2.

Al passo 4 abbiamo lasciato indefinito il metodo con cui definiremo la nuova politica in base ai nuovi valori di stati e azioni. Il metodo più naturale è di creare una politica *greedy* che in ogni stato segue sempre quella che al momento è stimata come la migliore azione. Se nello stato *s* il valore di *Q(s, a)* raggiunge il valore massimo per l'azione a^*, allora poniamo

$$\pi(s, a^*) = 1$$

$$\pi(s, a_i) = 0 \ \forall a_i \neq a^*$$

Fatto questo per ogni stato, la nuova politica è definita[81].

Questo metodo non tiene conto del *dilemma dell'esplorazione*. Tale comportamento è razionale se abbiamo già una capacità predittiva che giudichiamo sufficientemente affidabile e non avvertiamo l'esigenza di apprendere ancora, in altri termini non cerchiamo nuova informazione. Ma i MCM sono appunto metodi di apprendimento in cui non si ha un modello affidabile del sistema e si cerca invece di imparare continuamente. Ne consegue in modo naturale che non useremo politiche deterministiche, benché ciò sia lecito, ma politiche stocastiche, le quali lasciano spazio per la scelta di azioni che non appaiono ottimali in un certo stato ma che per qualche motivo sembrano degne di ulteriore apprendimento, sulle quali vogliamo quindi raccogliere ancora informazione. Riparleremo di questo tema in seguito.

---

[81] Essendo greedy, questa politica è deterministica. Non necessariamente vale il viceversa. Una politica potrebbe imporre una sola azione anche se questa non è quella stimata migliore.

Il nostro schema lascia aperto anche il problema della *generalizzazione dell'informazione*. Se lavoriamo con dati reali e non simulati, andremo molto spesso incontro al problema di avere un numero di episodi sperimentali insufficiente a stimare in modo affidabile gli stati, le azioni e le politiche. Più numerosi sono stati e azioni, più è elevato il bisogno di informazione. Questo problema è legato al dilemma del'esplorazione, ma non si esaurisce in questo, perché anche l'esplorazione può non essere sufficiente se i dati sperimentali sono comunque troppo pochi. Si avverte quindi l'esigenza di apprendere su certi stati, azioni e politiche anche quando essi non sono stati effettivamente incontrati, oppure lo sono stati troppo poco. Anche questo tema sarà trattato in seguito.

I MCM condividono diverse idee della DP: essi calcolano le stesse funzioni di valore, e raggiungono le politiche ottime con lo stesso metodo iterativo di valutazione-miglioramento. Presentano diversi vantaggi:

1. Apprendono *on*-line e non hanno bisogno di un modello iniziale del sistema; spesso nelle applicazioni la situazione è proprio questa.
2. Possono usare dati simulati; frequentemente è possibile generare tali dati anche senza avere un modello della distribuzione che li genera.
3. Rendo possibile focalizzare l'attenzione solo su alcuni stati e valutarli accuratamente, anche senza spendere risorse per valutare così accuratamente stati meno interessanti.
4. Sono meno sensibili alla violazione della proprietà di Markov (il motivo è che non fanno *bootstrapping*, cioè non costruiscono valutazioni su altre valutazioni).

Si può capire come questi metodi siano apprezzati nelle applicazioni, quando il rigore della DP risulta difficile da raggiungere, e in certa misura anche sacrificabile.

7.5 Apprendimento per differenze temporali

L'apprendimento per differenze temporali, usualmente chiamato *TD Learning*, è una combinazione di idee dei MCM e della DP: come i primi ha *l'apprendimento online* senza necessità di un modello del sistema, come la seconda fa *bootstrapping* e basa le valutazioni su altre valutazioni.

Come nei MCM, generiamo una serie di approssimazioni al valore $V^{\pi}(s)$ di uno stato sotto una certa politica. Se un certo stato s_t è visitato durante un episodio di apprendimento, entrambi i metodi ne aggiornano la stima $V^{\pi}(s_t)$. Nei MCM l'aggiornamento è effettuato solo a episodio terminato, usando R_t come obiettivo per la correzione della stima. Nei metodi TD, invece, la correzione della stima avviene già al passo $t + 1$, usando non il "vero" obiettivo R_t ma una sua approssimazione di un solo passo, quindi il flusso r_t:

$$V(s_t) \leftarrow V(s_t) + \alpha[r_t + \gamma V(s_{t+1}) - V(s_t)]$$

Anche qui, come nei MCM, si usa lo schema generale

$$Nuova\ stima \leftarrow Vecchia\ stima + Passo[Obiettivo - Vecchia\ stima]$$

È interessante notare che anche la DP si basa su una regola di questa forma, precisamente

$$V(s_t) \leftarrow V(s_t) + \alpha[r_t + \gamma V^{\pi}(s_{t+1})]$$

Anche se nei metodi TD l'obiettivo è definito diversamente, perché si utilizzano i dati di un solo passo, resta vero che esso si basa sui dati osservati, come nei MCM. È però anche vero che per la stima si usano altre stime, come nella DP. Infatti la parte di R_t che non compare in r_t viene attribuita al valore $V(s_{t+1})$ che non sta in $V(s_t)$. Mentre il valore di stato calcolato dai MCM è una stima del vero valore perché la somma di flussi osservata è un campione di quella vera, quello calcolato dalla DP non è una stima in questo senso (si assume di conoscere la dinamica di sistema) ma nel senso che la stima si basa su un'altra stima, perché dovremmo usare $V^{\pi}(s_{t+1})$e invece usiamo la sua stima $V(s_{t+1})$. Nei metodi TD abbiamo una stima in entrambi i sensi: i flussi sono osservati a campione, e le stime si basano su altre stime.

I metodi TD sono apprezzati per le applicazioni per i motivi detti: apprendono online senza richiedere un modello preesistente (come invece fa la DP) e si implementano in modo incrementale ad ogni nuova osservazione, senza attendere la fine di un episodio di apprendimento (come invece fanno i MCM). Uno svantaggio è che i requisiti affinché il metodo converga ai veri valori da stimare sono più stretti che per gli altri due metodi. In particolare, si richiede che il parametro α sia abbastanza piccolo e che decresca con il tempo; questo per evitare troppe oscillazioni nella stima fra un passo e i successivi.

A questo punto abbiamo un metodo TD per la valutazione degli stati. Ora dobbiamo usarlo per la ricerca di una politica ottimale. Come visto per gli altri metodi, l'idea è di trasformare ad ogni iterazione la politica corrente π in una politica più *greedy*, che si concentra sull'azione migliore disponibile per ogni stato.

La regola per aggiornare i valori delle azioni è

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

La regola segue lo schema ormai noto. L'uso della differenza fra due valori successivi di Q motiva il nome del metodo: la differenza temporale fra due stime, siano di V o di Q, permette di giustificare l'approssimazione di un solo passo al valore R_t; il *bootstrapping* rende possibile l'apprendimento incrementale.

Esiste un altro algoritmo di tipo TD, noto come *Q-Learning*, che si è rivelato estremamente efficace e gode di grande popolarità. L'idea è di iterare sulla funzione Q per approssimare direttamente Q^* *indipendentemente dalla politica*. La regola iterativa è

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$$

La politica ha ancora effetto, perché decide quali sono le coppie stato-azione che sono visitate e modificate, ma l'aggiornamento non si basa sulla politica stessa.

7.6 Implementazione dei modelli

In questa sezione svolgiamo alcune considerazioni che possono servire da guida per la riflessione a beneficio di un progettista di sistemi decisionali automatizzati o parzialmente automatizzati. Vogliamo mettere a fuoco i problemi progettuali affinché nel prosieguo di questo lavoro si possa inquadrare meglio il significato e il ruolo dei concetti che introdurremo, ad anche apprezzarne la rilevanza pratica.

Nel corso dell'esposizione sono emerse differenze e similitudini fra i tre metodi (DP, MCM e TD) che devono essere tenute ben presenti da chi vuole implementare un sistema applicativo basato su un modello come il nostro, e quindi anche sulle idee del Computational Learning e in particolare del Reinforcement Learning. La tendenza della ricerca è verso una visione unificata dei tre metodi, a conferma che il progettista di un sistema di questo genere deve tenere presenti alcuni concetti fondamentali che vanno oltre i particolari metodi e algoritmi.

Un primo aspetto essenziale di un progetto applicativo è la disponibilità di un modello per la predizione della dinamica del sistema, cioè la funzione di generazione di flussi ρ e la funzione di transizione di stato σ.

In origine le idee dei processi decisionali markoviani (MDP) e del Reinforcement Learning furono in parte sviluppate in contesti molto artificiali. Non a caso i ricercatori di Reinforcement Learning usano volentieri giochi come il Backgammon o il Poker per sperimentare le loro idee: in questi sistemi totalmente artificiali la dinamica è prevedibile, e i problemi sono invece di tipo più matematico e computazionale, avendo a che fare con un numero enorme di stati e di azioni. Un altro campo applicativo è quello della robotica, dove viceversa si può realisticamente supporre di sapere poco o niente sull'ambiente in cui il decisore opera (in questo caso il decisore è un robot, e una sonda di esplorazione della superficie di Marte alle prese con un ambiente sconosciuto è un ottimo esempio). A questi due estremi si situano chiaramente i metodi che assumono modelli completi e accurati, come la DP, e metodi che apprendono a partire da una *tabula rasa*, come gli MCM e il TD.

I contesti applicativi economico-aziendali sono di regola intermedi fra questi due estremi, poiché il decisore ha in genere una certa nozione di quello che può attendersi, ma sa anche che le sue convinzioni sono soggette a rischi di smentite e di obsolescenza. In un ambiente complesso, che pur essendo in buona parte di origine umana non ha certo solo le caratteristiche di un manufatto, ma anche quelle di un sistema naturale emergente da micro-interazioni, in un tale ambiente è realistico attendersi che il decisore disponga di modelli parziali per estensione e accuratezza, sia per la complessità di momento in momento che per la tendenza al cambiamento continuo.

Il progettista di un sistema decisionale deve quindi tenere presente la possibilità di modelli misti, che uniscono concetti dei tre metodi canonici, eventualmente implementando due modelli diversi in due tempi diversi, uno per la fase *exploration* in cui il modello del sistema è assente o parziale e inaccurato, e uno per i tempi successivi, quelli della *exploitation*, quando è disponibile un modello più affidabile. La distinzione fra le due fasi non può essere assoluta e in entrambe le fasi ci deve essere un bilanciamento fra le due componenti. Tuttavia, ha senso usare metodi *model-free* come gli MCM o il TD in prima fase e poi passare a metodi *model-based* come la DP. Si tenga presente che un sistema DP con politiche stocastiche è comunque in grado di offrire una soluzione al problema del dilemma dell'esplorazione: si tratta di assegnare i valori $\pi(s, a)$ in base ai valori $Q(s, a)$ in un modo che sia *greedy* in giusta misura. Vedremo che esistono modi efficaci di fare questo.

Per il progettista è essenziale saper valutare la bontà dei propri modelli predittivi e capire quale può essere il loro impatto economico sul processo decisionale. Sarà quindi importante avere una visione economica-decisionale del Computational Lerning, della quale daremo una nostra interpretazione.

Una scelta progettuale importante è la grana temporale dell'apprendimento: incrementale ad ogni arrivo di nuove osservazioni, come nel TD, o a intervalli come nei MCM? In linea di principio, un apprendimento a intervalli può essere più affidabile, meno oscillante nelle sue stime, ma può perdere in efficacia perché i dati arrivano in ritardo. Molto dipende dalla frequenza con cui arrivano le osservazioni e dai tempi naturali del processo decisionale. Se le decisioni devono essere prese molto frequentemente, questo può spingere verso un sistema più incrementale, cosicché le decisioni sono prese su dati aggiornati. A maggior ragione se i dati sono pochi oppure se vanno soggetti a rapida obsolescenza perché il sistema da governare è molto dinamico.

Si pensi a una applicazione di *news recommendation* su web, l'ormai classico "articoli correlati dei quotidiani in edizione online. Uno stato è una collezione di statistiche sulle aperture e letture di articoli, una azione è la proposizione di 5-10 articoli scelti fra centinaia, un flusso è un conteggio di aperture e letture. Una *breaking news* può far sì che i consigli che erano interessanti pochi minuti fa diventino di colpo obsoleti. In questi contesti il tempo di aggiornamento naturale delle liste di articoli consigliati è di 5-10 minuti. Ma questo lasso di tempo non è del tutto sufficiente per rilevare correttamente i "flussi generati", che in questo caso si possono misurare in aperture e letture di articoli consigliati. Con questa grana temporale, il problema dell'attribuzione è delicato, perché non è chiaro se un lettore ha aperto un articolo perché suggerito o in modo spontaneo, né se dopo l’apertura lo ha veramente letto (il server web può rilevare se il lettore dopo l’apertura di un articolo ci si è soffermato almeno un minuto, altrimenti lo considera aperto ma non letto). In tale contesto operativo, è ragionevole tendere verso un sistema discretamente incrementale. Se invece vendiamo libri online, si possono attendere certamente tempi più lunghi, alleggerendo il carico sui sistemi informatici e consentendo una rilevazione dei dati più affidabile.

Interessante per il progettista degli algoritmi è lo schema iterativo che compare in tutti i tre metodi canonici. Per approssimare una “vera” funzione come V o Q, si genera una sequenza di stime che approssimano la funzione mediante lo schema

$$Nuova\ stima \leftarrow Vecchia\ stima + Passo[Obiettivo - Vecchia\ stima]$$

Abbiamo visto che i tre metodi hanno tre definizioni diverse di cosa sia l’obiettivo, perché lavorano su tre metodi di stima, per cui la distanza fra il dato osservato e la vecchia stima può essere imputata a un errore nella stima dei flussi o degli stati. Ma il concetto di base è comune: ogni nuova osservazione attrae la stima verso di sé e l’attrae tanto più fortemente quanto più è lontana. Il parametro passo di apprendimento serve a smorzare questi movimenti della stima per evitare che essa oscilli troppo, in un certo senso rincorrendo ora questo ora quel nuovo dato. Più è grande il parametro, più è volatile la stima; più è piccolo il parametro, più la stima diventa conservatrice, al limite immutabile se il passo tende a zero.

Nel progettare le regole di aggiornamento, occorre regolarsi sulle specificità del caso. Un aspetto è la complessità computazionale, che può diventare proibitiva se si scelgono male oggetti e tempi dell’aggiornamento. Occorre per esempio capire se r_t, il flusso immediato di un certo periodo, è un buono stimatore di R_t, il flusso totale successivo. Se lo è, ciò incoraggia a usare un metodo incrementale a episodi di apprendimento brevi, altrimenti si penserà a episodi lunghi. Quanto la cosa sia vera, lo si apprenderà raccogliendo statistiche e vedendo quanto i flussi stessi sono varianti fra uno stato e l’altro, e anche nel tempo in uno stesso stato.

Considerazioni analoghe valgono per l’altro aspetto dell’ottimizzazione, l’utilizzo delle predizioni sui valori al fine di migliorare le politiche. Anche qui si è visto uno schema iterativo tipico: si modifica una politica facendola diventare *più greedy* a favore delle azioni più promettenti di ciascuno stato.

La natura più o meno markoviana del processo decisionale è molto importante e dipende non solo dal fenomeno in sé, ma molto anche dalla progettazione degli stati stessi. Se il progettista include nello stato del modello non solo i dati che descrivono il presente “fisico” del fenomeno, ma anche dati memorizzati relativi a eventi passati, allora il processo diventa più markoviano semplicemente perché lo stato è esso stesso “più storico”. Il prezzo da pagare è una maggiore complessità sia computazionale che concettuale. Il progettista può anche scegliere metodi più robusti rispetto a deviazioni dalla proprietà markoviana. Per esempio, in genere gli MCM sono più tolleranti della DP.

Il grado di dinamicità, o viceversa di stazionarietà del processo è evidentemente molto importante, e ancora dobbiamo approfondire la questione trattando del dilemma dell’esplorazione. Il numero di stati e azioni è importante per capire se l’apprendimento e l’approssimazione possono avvenire in tempi utili, e qui entra in gioco la generalizzazione dell’informazione. Questi due temi sono fondamentali e dal loro trattamento dipende in larga misura l’efficacia di un sistema ispirato al nostro modello decisionale, specie se si è nella situazione descritta prima in cui l’informazione sul sistema ci permette di effettuare predizioni ma non in modo abbastanza affidabile.

Infine, il più immediato dei problemi pratici in molte applicazioni: i dati sono pochi e arrivano lentamente, il che di nuovo ci riporta ai due problemi cruciali: sfruttare al massimo l’informazione disponibili mediante generalizzazione, e ottenere rapidamente informazione utili mediante esplorazione.

8 GENERALIZZAZIONE DELL'INFORMAZIONE

In questo capitolo affronteremo un problema già comparso più volte: generalizzare l'informazione, nel senso di utilizzare l'informazione osservata su certi oggetti (stati, azioni) per formulare ragionevoli ipotesi su altri oggetti.

Nel contesto dei MDP la generalizzazione è spesso denominata *approssimazione di funzioni*, perché il problema è di generare stime sempre più vicine a una funzione incognita, che potrà essere il valore di uno stato *V*o il valore di una azione *Q*.

8.1 Apprendimento supervisionato di una funzione

Il nostro obiettivo è individuare una equazione $Y = f(X)$ che permetta di prevedere quale valore assumerà *Y* per ogni valore di *X* . Sappiamo che la funzione matematica *f* non coinciderà di regola con il reale legame fra le due variabili, ma il nostro obiettivo è di raggiungere una accuratezza soddisfacente, non assoluta. Per la precisione, non è necessario che la *f* sia una funzione matematica in forma esplicita, ma può anche essere un algoritmo che ha *X* come input e *Y* come output.

Per apprendere la funzione tramite addestramento supervisionato procediamo in questo modo:

4. Prendiamo un *training set* di dati della forma (x_i, y_i), coppie di manifestazioni correlate di *X* e *Y*.
5. Specifichiamo la forma della funzione o dell'algoritmo *f*.
6. Sottoponiamo il training set e la forma funzionale a un algoritmo approssimatore di funzioni, che restituisce una funzione o algoritmo per *f*.
7. Misuriamo l'accuratezza con cui *f* approssima i dati del training set, cioè il grado in cui i valori $f(x_i)$ "previsti"[82] dalla *f* sono effettivamente vicini ai valori osservati in passato y_i.
8. Misuriamo l'accuratezza della *f* anche su un *validation set* analogo al training set, ma contenente dati che non sono stati utilizzati per generare la *f*.
9. Se l'esito è giudicato soddisfacente, usiamo la *f* per effettuare previsioni su nuovi valori di *X*, per i quali non si consoce in anticipo il "vero" valore di *Y*.

---

[82] Stiamo usando la *f* per prevedere valori passati, il ch è legittimo perché si tratta di valori incogniti. Anziché di valori previsti, potremmo parlare di valori stimati, approssimati o anche spiegati, nel senso che una *f* accurata "spiega bene" i valori osservati in passato.

10. Nel tempo si aggiorna la *f* utilizzando nuovi dati osservati.

La forma funzionale è uno schema di funzione con alcuni parametri lasciati liberi. Il compito dell'algoritmo di apprendimento è fissare i migliori valori di questi parametri, migliori nel senso che la funzione approssimata è la più accurata possibile fra quelle della sua stessa forma. Per esempio, possiamo dire all'algoritmo di apprendimento che la *f* dovrà avere la forma lineare $aX + b$ (una retta nel piano cartesiano). L'algoritmo sceglierà i due valori dei parametri a e b tali che l'insieme delle coppie $(x_i, ax_i + b_i)$ sia il più possibile vicino ai valori (x_i, y_i) del training set. Si noti che il concetto di "vicino" resta ancora da definire.

Il validation set è distinto dal training set perché il nostro scopo è di far apprendere all'algoritmo di approssimazione qualcosa di utile su dati futuri. La previsione dei dati passati non ci interessa di per sé, ma come esercizio per acquisire abilità trasferibili nel futuro. Per fare un paragone, lo studente risolve esercizi a risposta a lui o lei nota per apprendere abilità: questi esercizi sono il training set. Poi sostiene un esame su esercizi la cui risposta gli è ignota, ma è nota all'esaminatore che così può valutare la sua capacità di generalizzare le competenze acquisite e trasferirle su nuovi problemi: questi esercizi sono il validation set. Infine lo studente esce dall'addestramento, entra nella vita professionale e applica le sue competenze su problemi a risposta ignota a tutti: questi problemi sono i nuovi dati che richiedono predizioni.

Esiste una grandissima varietà di tecniche di apprendimento supervisionato. La Statistica tradizionale ne offre alcune, come la regressione lineare e altre forme di regressione. Altre vengono dal settore del Computational Learning e dal Data Mining, che di quello è una branca con certe peculiarità di metodo e di obiettivi. Ne nominiamo alcune: reti neurali, algoritmi genetici, alberi di classificazione, metodi basati su casi, macchine a vettori supporto, reti bayesiane.

È comune parlare della *f* generata dall'apprendimento come di un *modello* della funzione reale.

8.2 Accuratezza e complessità dell’approssimazione di funzioni

Il criterio per misurare l’accuratezza dell’approssimazione di una funzione non è unico. Il criterio più noto e diffuso è il criterio della minimizzazione dell’errore quadratico medio (MSE), che consiste nel minimizzare

$$MSE = E\{(y_i - f(x_i)^2\}$$

dove E sta per il valore atteso della quantità in parentesi graffe. Se i valori x_i hanno tutti lo stesso peso in probabilità, allora il valore atteso è la tradizionale media aritmetica dei quadrati degli scarti fra i valori reali y_i e quelli previsti $f(x_i)$.

Non è affatto scontato che minimizzare MSE sia l’obiettivo dell’apprendimento e della approssimazione di *f*, e infatti esistono altri criteri. Per giustificarlo intuitivamente prendiamo un caso limite.

Supponiamo che la forma funzionale desiderata sia $f = c$, dove *c* è una costante. Vogliamo predire i valori che assumerà *Y* usando un unico numero, e cerchiamo il migliore *c* possibile. L’intuizione suggerisce che la migliore costante, la più rappresentativa dell’insieme di valori che può assumere *Y* è la media di *Y*, che infatti è il primo numero che ci viene in mente quando vogliamo sintetizzare l’informazione relativa a un insieme di numeri. È facile dimostrare che la media di *Y* è proprio il numero che minimizza MSE e che MSE è la ben nota varianza di *Y*. L’interesse di questa considerazione è che abbiamo scelto per *f* la più semplice forma funzionale possibile, una costante, e che in questo caso limite, *nel quale l’informazione su X non è inclusa*, MSE e il suo minimizzatore sono le due grandezze classiche e intuitive della statistica elementare.

Il caso limite è tale non solo nel senso formale di modello di minima complessità matematica, ma anche nel senso che è un modello che non utilizza affatto l’informazione su *X*. Se il training set è

$$\{(2,4),(3,6),(4,8)\}$$

costruendo un modello $f(x) = c$ ricordiamo solo che i valori della *Y* sono

$$\{4,6,8\}$$

e il miglior modello che possiamo generare è

$$f(x) = 6$$

La relazione *Y è il doppio di X*, che ha accuratezza perfetta nel predire il training set, non può essere recuperata dai solo valori di *Y*. Se quindi confrontiamo i due modelli

$$y = f(x) = 6$$

$$y = f(x) = 2x$$

è intuitivo che *la differenza fra le due accuratezze misura il valore dell'informazione relativa a X*. È come se noi avessimo questa tabella

X	**Y**
?	4
?	6
?	8

e dovessimo prevedere i futuri valori che *Y* assumerà. Possiamo acquistare i valori incogniti della *X*. Se non li acquistiamo, allora il meglio che ci resta è creare il modello non informato *y = 6*, e ci troviamo con *MSE = varianza di Y*. Se invece li acquistiamo, allora il possiamo creare il modello completamente informato *y = 2x* e ci troviamo con *MSE = 0*. Il guadagno di MSE è pari alla varianza e al domanda è: qual è il prezzo che siamo disposti a pagare per avere i dati mancanti?

Assumiamo dunque la minimizzazione di MSE come ragionevole criterio di accuratezza per i nostri modelli che approssimano funzioni. In concreto, noi avremo bisogno di un modello che predica $V^{\pi}(s)$ per gli stati del modello decisionale. Dopo t passi di approssimazione abbiamo una funzione V_t che stima V^{π}. Definiamo MSE_t come il valore atteso dello scarto quadratico fra la vera funzione e la sua approssimazione:

$$MSE_t = \sum_s \Pr(s)\,[V^{\pi}(s) - V_t(s)]^2$$

Si tratta del valore atteso rispetto agli stati, cioè la media pesata degli errori quadratici, dove i pesi sono le probabilità che un certo stato si presenti. Le probabilità potranno essere note grazie a un modello, oppure stimate empiricamente su campioni storici, oppure stimate a loro volta con qualche approssimazione. La pesatura fa sì che l'apprendimento cerchi di avere la massima accuratezza sugli stati che saranno incontrati più frequentemente, e sarà più tollerante all'errore sugli altri.

La funzione di approssimazione *f* avrà dei parametri, p.e. il nostro modello costante ne ha uno, *c*, mentre il modello lineare ne ha due, a e b. La indichiamo perciò come $f(\vec{\theta})$ mettendo in evidenza il vettore dei suoi parametri. Nel nostro caso, l'approssimazione V_t avrà un certo vettore di parametri e l'errore MSE_t dipenderà da questi parametri: per minimizzarlo dovremo scegliere i migliori. Quindi riscriviamo l'errore

$$MSE(\vec{\theta}_t) = \sum_s \Pr(s)\,[V^\pi(s) - V_t(s)]^2$$

Il problema dell'approssimazione è scegliere il miglior vettore di parametri possibile, per minimizzare l'errore.

La flessibilità di una funzione dipende in genere dal numero dei suoi parametri, e in pratica gli stati sono sempre molto più numerosi dei parametri. Possiamo dire, in termini economici, che la flessibilità del modello di approssimazione è una risorsa scarsa. Poiché, sempre parlando in generale, l'accuratezza dipende dalla flessibilità, ne consegue che abbiamo un trade-off tra accuratezza e flessibilità. A sua volta maggiore flessibilità significa maggiore complessità computazionale e di pensiero e in conclusione abbiamo un trade-off tra accuratezza e semplicità. Noi vorremmo da un modello entrambe le qualità, semplice e accurato, ma di regola c'è un conflitto fra le due esigenze e dobbiamo cercare un compromesso. Nella pratica, i progettisti di sistemi applicativi apprezzano molto la semplicità dei modelli. Un motivo importante è che in un sistema reale si deve sempre tenere conto di innumerevoli particolari che rovinano l'eleganza matematica e algoritmica del modello: dati mancanti, dati errati, vincoli idiosincratici di un'azienda, retaggi di sistemi risalenti a un altro contesto operativo. Questo "rumore" è più gestibile se il modello è a bassa complessità, con pochi parametri, facile da comprendere per un utente finale o uno sviluppatore software.

8.3 Selezione e generalizzazione di modelli

Il progettista di un modello di approssimazione di funzione deve in primo luogo selezionare una forma di modello, per esempio se vuole approssimare la funzione incognita con una costante, una funzione lineare, polinomiale, esponenziale o di qualsiasi forma. Assumiamo per semplicità che uno stato s sia rappresentato da un solo numero, per esempio una capacità, la massima quantità di prodotto vendibile lungo tutto il processo decisionale. La funzione $V^{\pi}(s)$ potrà essere approssimata da una di queste funzioni $V(s)$:

$$V(s) = c$$

$$V(s) = as + b$$

$$V(s) = as^2 + bs + c$$

$$V(s) = a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0$$

$$V(s) = a \cdot s^b$$

$$V(s) = a\ log(s) - b$$

La selezione della forma di modello richiede una comprensione della natura specifica del fenomeno che si vuole rappresentare e al tempo stesse delle esigenze operative in cui il modello sarà inquadrato, come il trade-off tra complessità e accuratezza di cui abbiamo detto. Esiste un *triplo trade-off* fra

- la complessità formale del modello (lineare ecc.),
- la numerosità dei dati di training,
- la capacità di generalizzazione su nuovi esempi

Al crescere della numerosità dei dati cresce anche la capacità di generalizzazione, nel senso che sui nuovi esempi il modello ha un errore più piccolo. Al crescere della complessità del modello cresce anche la capacità di generalizzazione, ma solo fino a un certo punto, oltre il quale comincia a decrescere. La capacità di generalizzazione di un modello troppo complesso può essere sostenuta dall'aumento della numerosità dei dati, ma solo fino a un certo punto.

Se il modello è troppo semplice o troppo complesso rispetto alla "vera" funzione che stiamo osservando a campione, ci troviamo davanti a due fenomeni detti *underfitting* e *overfitting*.

Se cerchiamo di adattare una retta

$$V(s) = a(s) + b$$

a una funzione che in realtà è cubica

$$V^{\pi}(\mathrm{s}) = 5\mathrm{s}^3 + 2\mathrm{s}^2 + 3\mathrm{s} - 18$$

l'errore, misurabile come MSE, resta inevitabilmente alto e non basta aumentare i dati di training, ma dobbiamo far crescere la complessità del modello. Non possiamo pretendere che una retta approssimi bene una parabola, un'iperbole, una curva sinusoide. In questi casi parliamo di *underfitting*.

Può accadere l'inverso, con una funzione complessa che deve adattarsi a dati semplici (generati da una funzione semplice). In questi casi possiamo trovarci con una curva molto oscillante, che rende conto dei dati perché compie giravolte che non è realistico aspettarsi nei fenomeni reali.

Ma soprattutto una funzione troppo complessa può apprendere molto bene da un campione poco rappresentativo della popolazione da cui è estratto, finendo per apprendere non solo la vera natura della funzione incognita, ma anche il *rumore* da cui il campione è affetto. Questo è il fenomeno noto come *overfitting*.

Vale la pena formalizzare meglio queste due patologie della generalizzazione.

Supponiamo di voler stimare il flusso r con una funzione dello stato $f(s)$. Abbiamo una serie di campionamenti $(s_1, r_1), \dots (s_M, r_M)$.

Definiamo

$$Errore\ quadratico\ atteso = E_s[(r(s) - f(s))^2$$

$$Rumore = E_s[(r(s) - E_s[r(s)])^2]$$

$$Errore\ quadratico = (E_s[r(s)] - f(s))^2$$

L'errore quadratico atteso (pesato sulle probabilità di incontrare i singoli stati) misura lo scarto fra la vera distribuzione del flusso r e il suo modello f. Il rumore è la varianza intrinseca di fattori non inclusi nel modello, che infatti non compare nell'espressione. L'errore quadratico esprime la deviazione del modello *f* dalla vera media della grandezza da approssimare *r* e dipende dal training set. Vale questa equazione

$$Errore\ quadratico\ atteso = Rumore + Errore\ quadratico$$

Abbiamo scomposto l'errore di approssimazione in una grandezza che dipende dal training set e una che non ne dipende. In questo modo possiamo esprimere il fatto che una approssimazione *f* può essere molto buona su un campione e molto cattiva su un altro.

Per esprimere la bontà del modello *f* analizziamo l'errore quadratico atteso al variare dei campioni.

$$Errore\ quadratico\ atteso = E_c[(E_s[r(s)] - f(s))^2]$$

$$Bias = (E_s[r(s)] - E_c[f(s)])^2$$

$$Varianza = E_c[(f(s) - E_c[f(s)])^2]$$

Il *bias* misura l'errore intrinseco del modello indipendente dal variare dei campioni, mentre la *varianza* misura la dispersione dei valori calcolati dal modello attorno alla loro media al variare dei campioni.

Non conosciamo la vera funzione *r(s)* né conosciamo la distribuzione del rumore, quindi non possiamo calcolare il vero bias e la vera varianza. Possiamo però stimarli raccogliendo *M* campioni, ciascuno con *N* coppie di valori stato-flusso.

Definiamo la media campionaria di *f(s)*:

$$\bar{f}(s) = \frac{1}{M}\sum_{i=1}^{M} f_i(s)$$

dove ogni f_i è generata tramite addestramento su un campione (tutte condividendo la forma funzionale scelta per *f*). L'espressione indica il valore medio della *f(s)* in ogni sua variante su ogni dato di addestramento.

Ora possiamo stimare empiricamente il bias e la varianza

$$Bias^2(f) = \frac{1}{N}\sum_{(s_k,r_k)} \left[\bar{f}(s_k) - r(s_k)\right]^2$$

$$Varianza(f) = \frac{1}{NM}\sum_{c_i}\sum_{(s_k.r_k)\in\in c_i}\left[f_i(s_k) - \bar{f}(s_k)\right]^2$$

Prendiamo un caso di modello di minima complessità

$$f_i(s) = 2$$

Questo modello non ha varianza, perché indipendentemente dallo stato tutti i flussi stimati sono uguali. Però ci si attende che il bias sia elevato, salvo naturalmente che i flussi siano davvero tutti prossimi a 2. La semplicità del modello, quindi la sua rigidità, fanno sì che sia "puntato male" ma non oscillante.

Ora prendiamo un modello appena più complesso, che invece di una costante qualsiasi usa la media dei flussi nel campione *i*-esimo

$$f_i(s) = \sum_{(s,.)\in c_i}\frac{r(s)}{N}$$

Ora il bias diminuisce, perché la media è la migliore costante predittiva, però la varianza aumenta, perché i diversi campioni hanno diverse medie e i valori di flusso oscillano (di nuovo, salvo casi particolari).

Se il modello diventa più complesso, per esempio un polinomio di grado elevato, piccoli cambiamenti nel training set provocano grandi oscillazioni nel polinomio e quindi la varianza aumenta. Di contro, il bias diminuisce perché il polinomio si adatta bene ai dati. La complessità del modello, e quindi la sua flessibilità, fanno sì che sia "puntato bene" ma molto oscillante.

Siamo in presenza del fondamentale *trade-off bias-varianza*. Il progettista cercherà il miglior bilanciamento possibile fra le due esigenze.

La presenza di un elevato bias indica che il nostro modello non ha una forma abbastanza flessibile: siamo in *underfitting*. La presenza di una elevata varianza indica il modello ha una forma troppo flessibile e apprende anche il rumore: siamo in *overfitting*. Se aumentiamo il numero di dati di training, allora i dati riescono meglio a vincolare il modello e ne riducono la varianza, riducendo quindi il rischio di overfitting.

Nella pratica per regolare la complessità del modello si ricorre alla procedura di *cross-validation*. Non si calcolano bias e varianza, ma l'errore totale. Si prendono diversi modelli candidati di diversi livelli di complessità e ognuno è addestrato sul training set e valutato sul validation set, misurando l'errore. Questo errore decresce al crescere della complessità, poi arresta il miglioramento oppure migliora molto poco, o talvolta anche riprende a crescere. Il punto di "gomito" indica il miglior livello di complessità.

Avendo a disposizione dati a sufficienza, questa è la procedura standard nella prassi progettuale. Esistono altri metodi, di aiuto quando si dispone di pochi dati.

8.4 La discesa del gradiente

Il più classico dei metodi di ottimizzazione è il metodo della *discesa del gradiente*, che i progettisti prendono sempre in considerazione come una delle prime scelte.

Prendiamo il caso in cui lo stato è rappresentato da un singolo numero, per esempio una capacità di vendita, la funzione da approssimare è V^{π} e il modello ha un solo parametro θ. Per esempio, il modello potrebbe essere $V(s) = \theta s$. La regola della discesa del gradiente avvicina la stima al valore obiettivo, come già visto in altri casi, ma stavolta l'ampiezza dell'aggiustamento dipende non soltanto dal passo di apprendimento α e dall'ampiezza dell'errore $V^{\pi}(s_t) - V_t(s_t)$ ma anche dalla velocità con cui la stima si avvicina al suo obiettivo. Esprimendo la velocità come derivata di funzione, la formula della discesa del gradiente è

$$\theta_{t+1} = \theta_t + \alpha[V^{\pi}(s_t) - V_t(s_t)] \frac{dV_t(s_t)}{d\theta}$$

Intuitivamente, essendo la derivata il limite del rapporto incrementale $\frac{variazione\ di\ V_t}{variazione\ di\ \theta}$, essa approssima il rapporto incrementale, se è piccolo. La formula deriva chiaramente dall'approssimazione di funzioni con lo sviluppo in serie di potenze, unitamente all'introduzione del passo di apprendimento per evitare le eccessive oscillazioni su ogni nuova osservazione e tenere conto in piccola parte di ciascuna osservazione.

Ad ogni passo il valore della stima viene modificato per diminuire l'errore. Il metodo converge verso gli ottimi locali, non quelli globali. Quando si arriva a un valore θ' per il quale ogni piccola variazione fa aumentare l'errore, il procedimento si arresta. È possibile che esista un altro valore θ'' che approssima meglio la funzione obiettivo V^{π}, ma il metodo procede a piccoli passi, e quindi non può trovare il secondo valore se questo non è in un piccolo intorno del primo.

Il problema della convergenza a minimi locali dell'errore di approssimazione è certamente importante e da tenere presente, ma la sua portata pratica non deve neanche essere sopravvalutata. Spesso nelle applicazioni il valore ottimo è unico. Nel nostro semplice esempio *stato = capacità di vendita*, con un solo parametro moltiplicativo, è realistico pensare che il parametro che dà il minimo locale dia anche il minimo globale. Se si teme che non sia così, si può ricorre a una *perturbazione* del parametro che di tanto in tanto lo fa saltare di un passo non piccolo (ecco un esempio di *esplorazione* per acquisire nuova informazione su nuove regioni di valori del parametro). Oppure, più semplicemente, si può innescare il procedimento partendo da diversi valori iniziali di θ e scegliere tra i risultati quello che dà il valore minimo.

Tornando alla regola di aggiornamento iterativo, poiché nei casi applicativi interessanti V^{π} sarà una funzione stocastica, non si conoscerà il suo vero valore in s_t, ma soltanto una sua osservazione campionaria v_t. La regola può essere applicata usando l'osservazione come stima del valore ignoto:

$$\theta_{t+1} = \theta_t + \alpha[v_t - V_t(s_t)]\frac{dV_t(s_t)}{d\theta}$$

Se l'osservazione è uno stimatore privo di bias, cioè se $E\{v_t\} = V^{\pi}(s_t)$ e α decresce nel tempo, il metodo converge ugualmente verso un minimo locale dell'errore.

Quanto visto si generalizza per modelli con più di un parametro. La forma vettoriale della regola di discesa del gradiente è

$$\vec{\theta}_{t+1} = \vec{\theta}_t + \alpha[v_t - V_t(s_t)]\nabla_{\vec{\theta}_t}\mathrm{V_t(s_t)}$$

dove

$$\nabla_{\vec{\theta}_t}\mathrm{V_t(s_t)} \stackrel{\text{def}}{=} \left(\frac{\partial \mathrm{V_t(s_t)}}{\partial \theta_t(1)}, \cdots, \frac{\partial \mathrm{V_t(s_t)}}{\partial \theta_t(n)}\right)^{\mathrm{T}}$$

è il vettore riga (T sta per la trasposizione) delle derivate parziali.

Nella versione a più dimensioni la modifica dei parametri avviene nella direzione più promettente per ridurre l'errore, quindi con variazioni maggiori per i parametri più influenti nell'approssimazione.

Come sempre nella pratica, il progettista prenderà in considerazione spesso l'uso di un modello lineare, in cui lo stato è un vettore di caratteristiche, per esempio capacità totali dei vari prodotti, quantità messe in vendita, orizzonte temporale, come abbiamo visto descrivendo il nostro modello decisionale. Sia $\vec{\varphi}_s = (\varphi_s(1), \cdots, \varphi_s(n))^T$ il vettore delle caratteristiche di stato; in un modello lineare avremo un vettore di parametri della stessa dimensione $\vec{\theta}_t = (\theta_t(1), \cdots, \theta_t(n))^T$ e la relazione

$$V_t(s) = \vec{\theta}_t^{\,T} \vec{\varphi}_s = \sum_{i=1}^{n} \theta_t(i)\, \varphi_s(i)$$

In questo caso il gradiente coincide con le caratteristiche dello stato

$$\nabla_{\vec{\theta}_t} V_t(s_t) = \vec{\varphi}_s$$

ed esiste un solo minimo, pertanto la convergenza al minimo locale è anche convergenza al minimo globale.

Nella pratica accade spesso che la derivata non sia a rigore definita perché i parametri sono grandezze discrete, poniamo numeri interi. La discesa del gradiente è allora applicabile approssimando al derivata con il rapporto incrementale.

Poniamo di avere due parametri e che dopo *t* iterazioni il vettore di stima sia

$$\vec{\theta}_t = (3,7)$$

La derivata parziale di *V* rispetto al primo parametro potrà essere sostituita dal rapporto incrementale

$$\frac{V^{\pi}(4,7) - V^{\pi}(3,7)}{4 - 3}$$

oppure da

$$\frac{V^{\pi}(2,7) - V^{\pi}(3,7)}{2 - 3}$$

e lo stesso per le altre componenti del vettore gradiente.

In pratica, se campioniamo la funzione di valore di uno stato *s* con due serie di parametri quasi uguali a meno di una piccola variazione in un solo parametro, abbiamo una stima della derivata parziale, che potrà essere utile per applicare il metodo della discesa del gradiente.

9 IL DILEMMA DELL'ESPLORAZIONE

In questo capitolo affronteremo il problema del trade-off tra sfruttamento ed esplorazione quando esiste un solo stato e quindi si deve ottimizzare soltanto la scelta dell'azione. Descriveremo i principali metodi per l'esplorazione di azioni alternative, quindi avanzeremo alcune proposte su possibili miglioramenti rispetto allo stato dell'arte.

Successivamente descriveremo il metodo normale per affrontare il dilemma dell'esplorazione quando si considerano gli stati e ne evidenzieremo i limiti. Nel capitolo sui metodi avanzati di esplorazione vedremo come affrontare tali limiti.

9.1 Il problema dei banditi

Dapprima tratteremo il problema del'esplorazione in un modello con un solo stato, quindi in realtà prescindendo dallo stato. Il problema così posto non è affatto privo di interesse, ed è anzi ancora un problema aperto per la ricerca, nella quale è conosciuto come *Many-Armed Bandit (MAB) problem*. Un many-armed bandit è una slot-machine con più leve: ogni leva ha una sua propria regola di distribuzione delle vincite, ignota al giocatore. Il compito del giocatore consiste nell'elaborare una strategia di gioco per massimizzare la vincita complessiva, bilanciando l'esigenza di tirare le leve più redditizie e quella di imparare quali siano queste leve.

Il problema dei MAB ha anche un rilevante interesse pratico. Nella pubblicità online i MAB forniscono una cornice in cui inquadrare la selezione dell'annuncio pubblicitario da proporre su una pagina web. Ogni azione del nostro modello corrisponde a un annuncio. L'unicità dello stato significa che non si tiene conto di informazioni di contesto, ad esempio la natura del sito su cui visualizzare l'annuncio, oppure le caratteristiche del visitatore online che vedrà l'annuncio.

Nelle sezioni successive affronteremo questo problema: date K azioni $a_1, \ldots, a_K$ e i rispettivi valori di performance economica stimati finora $Q(a_1), \ldots, Q(a_K)$, assegnare le probabilità $\pi(a_1), \ldots, \pi(a_K)$ con cui ciascuna azione può essere selezionata. Le funzioni Q e π hanno un solo argomento, l'azione, perché lo stato è una costante uguale per tutte le azioni. La funzione Q in questo contesto è più semplice che in quello a più stati, perché $Q(a_i)$ dipende solo dalle osservazioni nei casi in cui a_i è stata selezionata, non essendoci propagazione di informazione fra stati e azioni.

Nei termini del modello di MDP, vogliamo approssimare le funzioni Q^* e π^*. Nei termini del nostro modello, vogliamo trovare una politica che massimizzi la somma dei flussi scontati tenendo in conto il patrimonio informativo disponibile e risolvendo il problema del giusto investimento in esplorazione per ottimizzare tale patrimonio informativo.

Nel seguito il contesto operativo di riferimento sarà quello online con dati reali: ad ogni periodo la politica viene aggiornata in base ai flussi generati nel periodo precedente, che costituiscono il *feedback* della politica vigente. Un episodio può essere lungo una o molte selezioni. Se pensiamo all'applicazione alla pubblicità web, una selezione corrisponde all'erogazione di un annuncio, le selezioni possono essere decine di migliaia al secondo, il feedback può arrivare dopo molti secondi. È quindi realistico pensare a episodi che coinvolgono moltissime selezioni. Questo aspetto non è però usualmente trattato nella letteratura scientifica, benché sia un problema rilevante per le applicazioni.

9.2 Stima di azioni a singolo stato

Le politiche che andiamo a esaminare usano una stima del valore di una azione $Q(a)$. Il modo più naturale di stimare il valore di una azione è la media storica dei flussi da essa generata. Se fino al periodo t l'azione a è stata scelta t_a volte la stima sarà quindi

$$Q_t(a) = \frac{r_1 + r_2 + \cdots + r_{t_a}}{t_a}$$

L'implementazione naturale di questa regola consiste nel memorizzare tutti i flussi generati sinora e al tempo t fare la somma dei valori dei flussi e dividere per il loro numero. Un metodo più efficiente è di aggiornare incrementalmente il valore di Q. Se Q_k è la media dei primi k flussi generati dall'azione a e r_{k+1} è un nuovo flusso per il quale dobbiamo aggiornare la media, la regola iterativa è:

$$Q_{k+1} = Q_k + \frac{1}{k+1}[r_{k+1} - Q_k]$$

La relazione iterativa vale anche per $k = 0$:

$$Q_1 = Q_0 + [r_1 - Q_0]$$

Se $Q_0 = 0$ allora è semplicemente $Q_1 = r_1$, come è naturale. Ma assegnare a Q_0 un valore non nullo è una operazione che ha perfettamente senso, come vedremo parlando del principio di ottimismo.

La formula per la stima incrementale di Q è ancora una volta della forma ben nota

$$Nuova\ stima \leftarrow Vecchia\ stima + Passo[Obiettivo - Vecchia\ stima]$$

Il passo qui decresce nel il tempo nel modo più naturale, in base al numero di flussi generati. Ciò significa però che il tempo non è il numero di episodi del processo decisionale, ma il numero di episodi in cui questa azione è stata scelta, episodi che possono essere più o meno recenti al momento della valutazione di Q.

Se il contesto operativo del processo decisionale non è stazionario, allora la stima di Q può essere modificata per tenerne conto, definendo un passo decrescente con il tempo di riferimento dell'intero processo, non solo della singola azione. Usiamo quindi la formula più generale

$$Q_{k+1} = Q_k + \alpha[r_{k+1} - Q_k]$$

dove il parametro α è costante, con $0 < \alpha \leq 1$. In forma non iterativa la definizione diventa

$$(1-\alpha)^k Q_0 + \sum_{i=1}^{k} \alpha(1-\alpha)^{k-i} r_i$$

Questa somma è una media pesata dei flussi, compreso il flusso fittizio Q_0. Si può verificare infatti che la somma dei pesi è uguale a 1:

$$(1-\alpha)^k + \sum_{i=1}^{k} \alpha(1-\alpha)^{k-i} = 1$$

Questa media pesata assegna maggior valore all'informazione più recente e progressivamente fa scivolare nell'oblio l'informazione più lontana nel tempo. La similitudine formale con la formula del reddito o del capitale economico come valore attualizzato di flussi non è casuale.

Si può assegnare a α un valore variabile nel tempo $\alpha_k(a)$, per esempio il valore $\frac{1}{k}$ come nel caso della semplice media dei flussi. La convergenza in probabilità della serie è garantita sotto due condizioni:

$$\sum_{i=1}^{\infty} \alpha_k(a) = \infty \text{ e } \sum_{i=1}^{\infty} {\alpha_k}^2(a) < \infty$$

La prima condizione è necessaria per garantire che i passi siano abbastanza grandi da superare prima o poi le condizioni iniziali e le fluttuazioni casuali. La seconda condizione assicura invece che questi passi diventeranno abbastanza piccoli da non generare essi stessi fluttuazioni.

Queste condizioni sono soddisfatte dal caso della media semplice $\alpha_k(a) = \frac{1}{k}$, ma non per un parametro costante $\alpha_k(a) = \alpha$. Infatti, nel secondo caso la stima continua a oscillare per rispondere alle variazioni dell'ambiente. M a in un ambiente non stazionario ciò è proprio quello che vogliamo: in effetti, nelle applicazioni la convergenza non è in genere una proprietà davvero richiesta, anche se nella letteratura scientifica più teorica viene ampiamente indagata.

9.3 Politiche estreme: RANDOM e GREEDY

Partiamo da due politiche che si pongono ai due estremi del dilemma exploration-exploitation: la politica RANDOM che non tiene in alcun conto la qualità delle azioni Q e quindi lo sfruttamento, e la politica GREEDY che viceversa si interessa solo allo sfruttamento e ignora le esigenze informative.

La politica RANDOM sceglie una azione a caso, con probabilità uguale per tutte:

Politica RANDOM
Input Valori delle azioni $Q(a_i)$, per $i = 1, \ldots, K$
Iterazione Per ogni i: $\pi(a_i) \leftarrow \frac{1}{K}$

L'interesse di questa politica è in primo luogo come termine di confronto per valutare la qualità delle altre, in analogia con l'approssimazione di funzioni mediante il valore medio della variabile dipendente y vista in precedenza.

La politica GREEDY assegna probabilità 1 alla migliore azione, probabilità 0 a tutte le altre:

Politica GREEDY
Input Valori delle azioni $Q(a_i)$, per $i = 1, \ldots, K$
Inizializzazione Seleziona una volta ciascuna azione a_i.
Iterazione $a^* \leftarrow \arg\max_i Q(a_i)$ $\pi(a^*) \leftarrow 1$ Per ogni $a_i \neq a^*$: $\pi(a_i) \leftarrow 0$

La politica GREEDY cambia soltanto quando cambia l'azione *leader*, quella con il massimo valore di Q. Perché questo avvenga occorre che il leader peggiori la sua performance fino a scendere sotto la seconda azione in classifica. Le azioni non leader non possono variare la loro performance, quindi il sorpasso può avvenire solo dall'alto verso il basso, dal primo posto in classifica in giù, non viceversa. La conseguenza è che se una azione si trova al primo posto e mantiene la performance fino a quel momento, non è più possibile avere informazioni sulle altre.

Se, per esempio, abbiamo due azioni a_1 e a_2 con la prima che genera flussi di valore costante 5 e la seconda flussi uniformemente distribuiti fra 1 e 20, se al primo turno la seconda azione genera un valore fra 1 e 4, da quel momento in poi sarà scelta per sempre la prima azione, anche se la seconda genera un reddito atteso che vale il doppio.

Anche la politica GREEDY ha un valore come caso limite. Abbiamo visto che certi metodi iterativi di ottimizzazione per MDP generano politiche greedy e che questo ha senso. Però in quei casi una azione poteva cambiare valore Q anche se non era selezionata, per propagazione all'indietro dei valori di stati successivi. Inoltre, in quei casi una condizione per la convergenza alla funzione ottimale Q^* era che ogni coppia stato-azione fosse visitata periodicamente, mai abbandonata. Pertanto si richiedeva comunque una certa esplorazione.

La politica GREEDY è certamente ottimale quando si arriva all'ultimo periodo dell'intero processo, perché non ha più senso esplorare per accumulare informazione utile in futuro, e quindi la scelta migliore è proprio selezionare l'azione che fin qui si è comportata meglio. Si intuisce quindi che in un processo a durata finita qualsiasi politica capace di esplorazione dovrebbe progressivamente tendere verso GREEDY per raggiungerla all'ultimo periodo. Se invece il processo è di durata infinita o almeno illimitata, allora questa considerazione non vale più.

Se il processo è non-stazionario, con i valori delle azioni che cambiano nel tempo, l'insufficienza di questa politica diventa ancora più evidente, perché tende a funzionare in base a presupposti obsoleti. È evidente che se il processo non è stazionario allora l'esplorazione diventa ancora più importante.

9.4 Politica ε-GREEDY

Nella letteratura scientifica e nella pratica professionale, il più tradizionale metodo per affrontare il dilemma dell'esplorazione è la politica ε-GREEDY, che oscilla fra le due politiche RANDOM e GREEDY. Il parametro ε, compreso fra 0 e 1 è una probabilità: ad ogni nuova selezione lo si usa per decidere se seguire RANDOM o GREEDY. In pratica, si estrae un numero casuale con distribuzione uniforme fra 0 e 1; se il numero è minore o uguale a ε allora si segue RANDOM, altrimenti si segue GREEDY.

Politica ε-GREEDY
Input Valori delle azioni $Q(a_i)$, per $i = 1, \ldots, K$ *Parametri* Probabilità $0 \leq \varepsilon \leq 1$ di scegliere RANDOM anziché GREEDY
Inizializzazione

Seleziona una volta ciascuna azione a_i.

Iterazione
 	
Con probabilità ε: $\pi(a_i) \leftarrow \frac{1}{K}$ per ogni a_i, oppure
 	
con probabilità 1 – ε:
 	
{
 	
$a^* \leftarrow \arg\max_i Q(a_i)$
 	
$\pi(a^*) \leftarrow 1$
 	
Per ogni $a_i \neq a^*$: $\pi(a_i) \leftarrow 0$
 	
}

In un processo stazionario di durata illimitata questa politica non converge a quella ottimale π^*, poiché ad ogni selezione la probabilità di scegliere l'azione migliore è limitata superiormente da $(1-\varepsilon)+\frac{1}{K}$.

Il parametro ε è scelto dal progettista. Non ci sono criteri generali; anche se l'usanza è che ε vari dal 5% al 25%, non c'è nessun motivo razionale a giustificazione di questa regola empirica. Peraltro, anche questo intervallo di valori è tropo ampio per essere davvero una guida rassicurante in un contesto professionale.

Se il processo ha durata limitata viene naturale diminuire nel tempo il parametro ε. Avremo quindi una sequenza decrescente di valori ε_t che tende a zero quando t tende a T, la durata totale del processo. Definire questa sequenza è un compito ancora meno razionalizzabile della scelta di ε. Peraltro, questo accorgimento non è molto menzionato nella letteratura e nei forum online scientifici e professionali.

In definitiva, la scelta di ε viene fatta semplicemente ad aggiustamenti successivi, cambiandolo e giudicando intuitivamente se i risultati sono convincenti.

Detto questo, occorre rendere giustizia a questa politica dicendo che anche le altre politiche menzionate nella letteratura scientifica richiedono la calibrazione di uno ma anche più parametri, spesso meno intuitivi di ε.

Il parametro ε è una quota fissa dedicata all'esplorazione, indipendentemente dalla configurazione dei $Q(a_i)$. Tale politica di controllo del processo è palesemente semplicistica, eppure, per quanto sorprendente, è diffusissima non solo nelle applicazioni ma anche nella letteratura scientifica. Anche i metodi sofisticati che abbiamo visto in precedenza, quando devono selezionare una azione in modo stocastico, di regola lo fanno proprio con la politica ε-GREEDY.

Il motivo della diffusione di una politica così primitiva è che, sorprendentemente, progettarne una che dia sistematicamente risultati migliori, senza doverla costruire ad hoc per una applicazione, è un compito molto più difficile di quanto può sembrare, sia dal punto di vista algoritmico che da quello tecnico di sviluppo software. I progettisti tendono quindi a risolvere in modo semplice e poco costoso il problema della politica di selezione delle azioni, affidandosi a uno strumento largamente condiviso, facilissimo da implementare e che dopotutto ha un parametro unico e accettabilmente intuitivo.

L'idea comunemente diffusa è che la politica ε-GREEDY sia una specie di "buon piano B per tutte le stagioni", mai eccellente ma raramente deludente. Il sottinteso è che progettare soluzioni migliori è un compito abbastanza ingrato e che offre poche garanzie di successo. In questi ultimi anni, con l'intensificarsi della ricerca scientifica motivata dal business della pubblicità online e l'apparizione di nuovi metodi, tale giudizio conservativo ha cominciato ad indebolirsi. Resta comunque vero che ε-GREEDY è ancora un benchmark con cui confrontare la progettazione di nuove politiche di controllo.

9.5 Politica SOFTMAX

La politica SOFTMAX ha una nobile discendenza scientifica, ispirandosi alla *macchina di Boltzmann*[83]. L'idea è approssimare la politica Greedy mantenendo però una certa chance di

[83] Spiegare perché.

esplorazione anche per le azioni non leader, che hanno sempre una probabilità non nulla di essere selezionate[84].

Politica SOFTMAX
Input Valori delle azioni $Q(a_i)$, per $i = 1, \ldots, K$. *Parametri* Temperatura $\tau > 0$ per regolare la distanza da RANDOM e GREEDY.
Inizializzazione Seleziona una volta ciascuna azione a_i.
Iterazione Per ogni a_i: $\pi(a_i) \leftarrow \frac{e^{\frac{Q(a_i)}{\tau}}}{\sum_{j=1}^{K} e^{\frac{Q(a_j)}{\tau}}}$

La probabilità di una azione di essere selezionata è proporzionale all'esponenziale del suo valore diviso per il parametro di *temperatura* (vedremo tra poco perché è così chiamato). Il denominatore della frazione è solo un fattore di normalizzazione che assicura che la somme delle probabilità sia 1.

La funzione esponenziale serve a esaltare le differenze fra i valori delle azioni. Non avrebbe infatti molto senso distribuire le probabilità di essere selezionate in modo proporzionale al valore, perché l'azione che interessa è solo la migliore, per quanto sia incognita. Con *K* azioni di valore non troppo diseguale si finirebbe per avere un limite superiore di probabilità di indovinare la migliore azione non molto superiore a *1/K*, limite inaccettabile.

Il parametro τ regola la misura in cui l'esponenziale esalta queste differenze di valori.

---

[84] Da qui il nome SOFTMAX: si sceglie l'azione con il massimo valore, ma in modo soft, non assoluto come in GREEDY.

Se τ tende a 0, allora i $Q(a_i)/\tau$ e con loro gli esponenziali tendono a ∞, ma i valori $Q(a_i)$ più grandi lo fanno molto più rapidamente degli altri, per effetto dell'esponenziale. Perciò *π(azione leader)* tende a 1, mentre le altre $\pi(a_i)$ tendono a zero. In sintesi, SOFTMAX tende a GREEDY.

Se τ tende a ∞,, allora i $Q(a_i)/\tau$ tendono a 0 e numeratori delle frazioni tendono a 1. Perciò tutte le $\pi(a_i)$ tendono a $1/K$. In sintesi, SOFTMAX tende a RANDOM.

Il nome di "temperatura" ha origini storiche, che derivano dall'uso di formule simili per descrivere fenomeni fisici in cui un sistema inizialmente caldo e caotico si stabilizza mentre si raffredda. La politica SOFTMAX viene utilizzata con una temperatura τ inizialmente elevata, quindi con poca differenziazione fra azioni ed elevata esplorazione. Con il tempo, la temperatura scende e si privilegiano sempre di più le azioni che si stimano migliori, come intuitivamente è giustificato.

Certamente SOFTMAX appare più raffinata rispetto a ε-GREEDY, per l'eleganza formale, l'unicità del criterio e la flessibilità offerta dal parametro τ rispetto a ε. Nella pratica scientifica e industriale, risulta abbastanza usata ma non quanto a ε-GREEDY. Il motivo è la difficoltà di regolare il parametro. Non solo non è semplice raffigurarsi mentalmente la curva esponenziale, ma soprattutto per scegliere valori significativi di τ occorre avere una stima dell'ordine di grandezza dei valori di *Q* che ci si può aspettare. Alcuni confronti in letteratura sembrano indicare performance poco soddisfacenti di questa politica, ma si tratta di prove lungi dall'essere risolutive.

9.6 Politiche REFERENCE VALUE

Le politiche viste in precedenza fanno dipendere la probabilità $\pi(a)$ che una azione venga scelta dal valore $Q(a)$ dell'azione stessa. La politica REFERENCE VALUE si basa su un concetto diverso. Si fissa un valore di riferimento che in qualche senso è un valore "giusto" per un flusso, e si giudicano "grandi" i flussi maggiori di esso e "piccoli" quelli minori. Quando una azione genera un flusso grande, la sua probabilità di essere selezionata in futuro aumenta, tanto più quanto il flusso è grande. Analogamente, se il flusso è piccolo la sua probabilità di essere selezionata diminuisce.

Anziché il valore di azione $Q_t(a)$ usiamo il concetto di *preferenza* per una azione $p_t(a)$. Le preferenze *p* possono essere usate proprio come i valori *Q* nei metodi ε-GREEDY o SOFTMAX. Per esempio, la probabilità di scegliere l'azione *a* al tempo *t* con una politica SOFTMAX sarà

$$\pi(a) \leftarrow \frac{e^{\frac{p_t(a)}{\tau}}}{\sum_{j=1}^{K} e^{\frac{p_t(a)}{\tau}}}$$

Come si vede, non è propriamente una politica ma piuttosto una famiglia di politiche che utilizzano le preferenze anziché i valori di azione.

Una politica di tipo REFERENCE VALUE è caratterizzata da un valore di riferimento $\bar{r}$ per i flussi, valore che è unico per tute le azioni. Questo valore può essere prefissato o variabile con il tempo. Una scelta naturale è la media pesata nel tempo dei flussi generati:

$$\bar{r}_{t+1} = \bar{r}_t + \alpha[r_t - \bar{r}_t]$$

con l'uso del ben noto parametro di passo di apprendimento $0 < \alpha \leq 1$.

Fissato in un modo qualsiasi il valore di riferimento, la regola di aggiornamento delle preferenze sarà ancora una volta quella target-errore:

$$p_{t+1}(a_t) = p_t(a_t) + \beta[r_t - \bar{r}_t]$$

dove il passo di apprendimento è $\beta > 0$.

I valori iniziali $\overline{r_0}$ e $\overline{p_0}$ possono essere fissati arbitrariamente, anche se la scelta più naturale è 0.

Le politiche di questo tipo possono essere interessanti laddove si disponga di un concetto di flusso normale o flusso obiettivo; in tali casi il valore di riferimento non è una media dei flussi, ma un valore a priori. Oppure possono essere un metodo di apprendimento in sé, quando il valore di riferimento è calcolato empiricamente in corso di processo decisionale.

9.7 Politiche PURSUIT

Le politiche della famiglia PURSUIT utilizzano sia i valori delle azioni Q sia le preferenze per le azioni p come i metodi del valore di riferimento. L'idea è di aumentare la probabilità di selezione dell'azione leader del momento, quella che ha il miglior valore:

$$a^*{}_{t+1} = \arg\max_a Q_{t+1}(a)$$

I valori Q delle azioni possono essere calcolati con uno dei metodi già visti, per esempio con una media pesata nel tempo dei flussi osservati.

La regola iterativa di aggiustamento della probabilità di selezione π è

$$\pi_{t+1}(a^*{}_{t+1}) = \pi_t(a^*{}_{t+1}) + \beta[1 - \pi_t(a^*{}_{t+1})]$$

La probabilità di selezione dell'azione greedy si sposta verso 1 di una frazione β della distanza attuale, mentre per tutte le azioni $a \neq a^*{}_{t+1}$ la probabilità di selezione diminuisce verso 0:

$$\pi_{t+1}(a) = \pi_t(a) + \beta[01 - \pi_t(a)]$$

e la somma dei π resta 1.

9.8 Il principio di ottimismo

I metodi fin qui visti cercano di risolvere il dilemma dell'esplorazione favorendo l'azione leader, come la politica GREEDY, ma al tempo stesso assegnando un "credito" alle altre azioni, affinché esse possano avere una chance di essere selezionate e forse rivelarsi migliori della leader. Però questo credito non ha una rappresentazione esplicita, è invece il risultato implicito di un meccanismo. Non si può quindi propriamente scindere la "desiderabilità" di una azione in una componente economica in senso stretto derivante dai flussi e una componente derivante dall'informazione, o se si preferisce dall'incertezza.

Ora esploreremo modi per dare una rappresentazione più esplicita del valore dell'informazione nel quadro del problema a singolo stato.

Il primo è il *principio di ottimismo*: si sovrastima volutamente il valore delle azioni meno esplorate, per le quali c'è maggiore incertezza sul valore vero. Non è ancora un vero e proprio assegnamento di valore informativo, perché la sovrastima è sul valore economico, però l'informazione, o se si vuole l'incertezza, è il parametro che regola la sovrastima, almeno in parte.

Il modo più semplice di applicare il principio di ottimismo è l'assegnazione di grandi valori iniziali nei metodi visti fin qui. Consideriamo la stima di *Q(a)* con la relazione iterativa

$$Q_{k+1} = Q_k + \alpha[r_{k+1} - Q_k]$$

che equivale alla formula non iterativa

$$(1-\alpha)^k Q_0 + \sum_{i=1}^{k} \alpha(1-\alpha)^{k-i} r_i$$

Se assegniamo un valore positivo a Q_0 diamo un vantaggio alle azioni meno esplorate, perché l'effetto di sovrastima si sconta a tasso $(1 - \alpha)$ ad ogni selezione dell'azione. In questo modo otteniamo un impulso all'esplorazione che si affievolisce nel tempo, lasciando che ogni azione abbastanza esplorata sia valutata in base ai suoi risultati effettivi.

Vediamo un esempio in un contesto concreto, tratto dalla pubblicità online.

Quando un visitatore apre una pagina web, il sito sceglie un annuncio pubblicitario fra molti da visualizzare. L'obiettivo è prendere più click possibile. Ogni azione corrisponde a uno degli annunci candidati alla selezione. Il flusso r_t vale 1 se c'è il click e 0 se non c'è. Il valore *Q(a)* è una stima del tasso medio di click sull'annuncio *a*.

Un aspetto critico di questo ambito applicativo è il basso tasso di click: uno su cento visualizzazioni è già una buona media. Quasi tutti gli r_t valgono 0 e occorrono molti tentativi prima di avere informazione significativa sul valore di una azione, cioè sul tasso di click di un annuncio. Peraltro, le visualizzazioni di una campagna pubblicitaria multi-annuncio si misurano in milioni.

In questo contesto una implementazione realistica del metodo del valore iniziale ottimistico potrebbe essere il seguente. Ad ogni azione *a* si attribuiscono al tempo zero 10.000 visualizzazioni e e 200 click fittizi. Perciò $Q_0(a) = 0{,}02$ che è un tasso di click ottimistico, doppio dello standard. La formula iterativa diventa

$$(1-\alpha)^k(0{,}02) + \sum_{i=1}^{k} \alpha(1-\alpha)^{k-i}(r_i + 200)$$

Il parametro 10.000 è presente nella formula, essendo implicito nei due numeri 0,02 e 200.

Il metodo del valore iniziale ottimistico è troppo semplice per rappresentare una soluzione generale al problema del dilemma sfruttamento-esplorazione. In primo luogo, il suo impatto è temporaneo e quindi adatto solo a un processo stazionario. Nell'esempio della pubblicità, l'esperienza dice che di regola gli annunci hanno un ciclo di vita in due fasi: nella prima hanno un tasso di click abbastanza costante, nella seconda declinano, ma non tutti con la stessa velocità. Si tratta di un processo prima quasi stazionario e poi decisamente non stazionario. Il valore iniziale ottimistico ha un effetto sulla prima fase, ma non più sulla seconda, perché quando gli annunci cominciano a variare il loro tasso ormai tutte le azioni sono così esplorate da avere scontato quasi tutto il valore iniziale. Perciò al momento in cui i tassi di click degli annunci (i veri valori di *Q*) cominciano a variare ormai il sistema ha perso l'impulso all'esplorazione di cui avrebbe bisogno. Se invece l'impulso del valore iniziale fosse ancora efficace nella seconda fase, ciò significherebbe che avrebbe falsato la prima fase, imponendo una esplorazione eccessiva.

Anche per un processo stazionario, dove il metodo è più applicabile, la scelta del valore iniziale da assegnare al parametro *Q(0)* non è intuitiva: occorre conoscere piuttosto bene il fenomeno e i valori normali per *Q*. Ciò non toglie che nelle applicazioni il metodo abbia spesso mostrato una sua utilità.

Il metodo del valore iniziale positivo non ha la sua unica motivazione come tecnica per implementare il principio di ottimismo. Se invece di scegliere un valore iniziale artificialmente elevato, scegliamo un valore realistico, possiamo accelerare l'apprendimento di *Q*. Si pensi all'esempio della pubblicità online. Se diamo un valore nullo ai $Q_{t=0}(a)$, pochi click presi ai primi tentativi possono dare un grosso vantaggio dilunga durata a un annuncio *a* che in realtà non è migliore dei suoi annunci competitori. Questo perché il rapporto fra *Q(a)* e un *Q(a')* può essere di 3 click a 1 su 1.000 visualizzazioni ciascuno. Ma se assegniamo a tutti gli annunci un valore iniziale, l'apprendimento dà un peso più realistico ai click ed è meno soggetto ad essere ingannato da casi fortuiti.

Si può anche pensare di assegnare valori iniziali diversi alle diverse azioni. Questo si traduce in indicazioni al decisore su quali ci si aspetta che siano le azioni migliori. Inizialmente il decisore tende a seguire quelle azioni, salvo cambiare idea se i risultati empirici contraddicono i presupposti.

Naturalmente, mentre un valore iniziale ottimistico favorisce l'esplorazione, uno pessimistico la ostacola, mentre una discriminazione iniziale fra le azioni aderente al vero accelera l'apprendimento, una lontana dai veri valori lo rallenta. Per questo una buona regolazione del valore iniziale fittizio richiede conoscenza del fenomeno. Per inciso, si noti che c'è una utilità sia nello scegliere un valore ottimistico (più esplorazione) che uno realistico (convergenza più rapida ai veri valori), mentre non c'è convenienza ad essere pessimisti.

Un modo più promettente di applicare il principio di ottimismo è di scomporre il valore di una azione in due valori distinti, uno relativo ai flussi e uno alla esplorazione:

$$Q(a) = Q_r(a) + Q_e(a)$$

Un metodo di questo genere sarebbe matematicamente più trattabile di quelli visti e offrirebbe più flessibilità quando implementato in un sistema software. Andiamo quindi a esaminare solzuioni di questo genere.

9.9 Politica INTERVAL ESTIMATION

Un modo di applicare il principio di ottimismo è di stimare il valore di una azione come *il più grande valore che, in base all'informazione disponibile, potrebbe assumere con una ragionevole verosimiglianza*.

La politica INTERVAL ESTIMATION[85] applica questa idea basandosi sul classico metodo di stima di intervalli di confidenza. Supponiamo per il momento che i flussi abbiano una distribuzione stocastica normale con media μ e deviazione standard σ. Dalla statistica di base sappiamo che con probabilità 95% la media osservata *m* di un campione di *n* flussi apparterrà a questo intervallo

[85] La nostra presentazione si basa sulle idee di [Kaelbling 1993].

$$\mu - 1{,}96\frac{\sigma}{\sqrt{n}} \leq m \leq \mu + 1{,}96\frac{\sigma}{\sqrt{n}}$$

Il moltiplicatore 1,96 è determinato dalla scelta del livello di confidenza 95%; se scegliessimo il livello di confidenza del 99% allora il moltiplicatore sarebbe 2,58, mentre per il livello di confidenza 80% il moltiplicatore sarebbe 1,28. Questi moltiplciatori si possono ricavare da qualunque testo di base di statistica.

Per l'applicazione al nostro problema stimiamo media e deviazione standard della distribuzione dei flussi, che sono parametri ignoti, usando il campione dei flussi osservati fino a questo momento durante il processo (inclusi eventuali flussi fittizi che volessimo introdurre come valori iniziali). La stima della media $\hat{\mu}$ è la media del campione $\frac{1}{n}\sum_i r_i$, mentre la stima della deviazione standard $\hat{\sigma}$ è la deviazione standard del campione s con la correzione $s\sqrt{\frac{n}{n-1}}$, che nelle applicazioni è di regola irrilevante.

Se abbiamo un livello di confidenza del 95%, allora solo il 2,5% dei campioni avrà media maggiore dell'estremo superiore dell'intervallo di confidenza (un altro 2,5% cadrà sotto l'estremo inferiore). Pertanto il 97,5% dei campioni avrà media minore o uguale al limite superiore $\hat{\mu} + 1{,}96\frac{\sigma}{\sqrt{n}}$. Usiamo questo limite superiore come valore ottimistico della media del campione di flussi, intendendo che con probabilità 97,5% la vera media sarà minore o uguale a tale valore. Se sostituiamo 1,96 con 1,28 allora questa probabilità diventa il 90% (80% più metà del 20% mancante, come detto prima).

In generale useremo come valore ottimistico

$$\hat{\mu} + \alpha\frac{\sigma}{\sqrt{n}}$$

Dove α è un parametro che dovremo fissare in funzione del nostro livello di confidenza desiderato, diciamo del nostro livello di ottimismo.

Con questo metodo abbiamo realizzato lo scopo di dividere in due il valore di una azione

$$Q(a) = Q_r(a) + Q_e(a)$$

ponendo

$$Q_r(a) = \sum_i r_i(a)$$

e

$$Q_e(a) = \alpha \frac{deviazione\ standard\ dei\ flussi\ di\ a}{\sqrt{numero\ dei\ flussi\ di\ a}}$$

Questa formula è soggetta a complicazioni se i flussi dell'azione a sono soggetti a trasformazioni quali lo sconto in base alla lontananza nel tempo. Prescindendo da questo, il principio è chiaro: si prende il più alto valore verosimile per l'azione, dove il significato di verosimile dipende dal parametro α, che ha la funzione di regolare l'esplorazione.

Il valore di esplorazione di una azione è maggiore se il numeratore è maggiore, quindi se i flussi sono più dispersi e la loro incertezza è maggiore. Abbiamo bisogno di più informazione per ridurre tale incertezza, una nuova esplorazione di questa azione è più utile di una nuova esplorazione di una azione che genera flussi meno dispersi. In altre parole, è più utile esplorare laddove la stima del valore medio dei flussi è meno affidabile piuttosto che dove ragionevolmente la stima basata sulle osservazioni è già accurata. Il valore di esplorazione cresce anche al decrescere del denominatore, quindi è maggiore per le azioni fin qui poco esplorate. In sintesi, Q_e è maggiore per le azioni i cui flussi osservati sono pochi e dispersi.

Abbiamo ipotizzato che la distribuzione stocastica dei flussi fosse normale. Questa ipotesi è spesso ragionevole, e comunque tende ad essere realistica al crescere delle dimensioni dei campioni. Prendiamo come esempio ancora la pubblicità online, quando il valore di una azione è la probabilità che il relativo annuncio riceva un click quando visualizzato. Il singolo evento non segue la distribuzione normale, ma è un evento binario, con flusso di 1 o 0 click, e quindi è di tipo bernoulliano. La sequenza di eventi segue quindi la distribuzione binomiale, e la stima della vera media (il vero tasso di click) richiederebbe l'uso della complessa distribuzione Beta. Però la Beta può essere approssimata con la Normale se i campioni sono grandi abbastanza, e questo è proprio il caso pratico consueto. Pertanto, l'uso dei moltiplicatori validi per la Normale è legittimo. Si consideri, del resto, che il parametro α si può fissare liberamente, e che se il suo significato intuitivo non è esattamente quello teorico ciò non comporta alcun errore, solo una approssimazione.

Politica* I*NTERVAL* E*STIMATION
Input Per ogni azione a_i: media osservata μ_i, deviazione standard osservata σ_i e numero n_i dei flussi di a_i. *Parametri* Parametro di esplorazione $0 < \alpha < 1$ per regolare il livello di confidenza della stima superiore della vera media dei flussi di una azione.
Inizializzazione Seleziona una volta ciascuna azione a_i.

Iterazione
 	
Per ogni a_i:
 	
{
 	
$Q_r(a_i) \leftarrow \mu_i$
 	
$Q_e(a_i) = \alpha \frac{\sigma_i}{n_i}$
 	
$Q(a_i) = Q_r(a_i) + Q_e(a_i)$
 	
}
 	
$a^* = \arg\max_a Q(a)$
 	
$\pi(a^*) = 1$
 	
$\pi(a) = 0 \; per \; a \neq a^*$

Il criterio di scelta è semplicemente di selezionare sempre l'azione che ha il massimo valore ottimistico. Pertanto la politica INTERVAL ESTIMATION è una GREEDY in cui il valore empirico *Q(a)* è sostituito da quello ottimistico.

Questo metodo è decisamente interessante, perché il suo significato è ben definito matematicamente e perché isola i due tipi di valore, economico immediato e informativo (essendo il valore informativo un valore economico incerto e futuro).

9.10 Politica UPPER CONFIDENCE BOUND

L'idea alla base della politica INTERVAL ESTIMATION è stata ripresa nella letteratura scientifica in molte versioni basate sull'idea usare come valore di una azione un limite superiore della stima della media dei suoi flussi. Un primo metodo[86] usa come limite superiore

$$Q(a_i) = \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ij} + \sqrt{\frac{2\log N}{n_i}}$$

dove il primo addendo è il valore $Q_r(a_i)$ calcolato come la media degli n_i flussi generati dall'azione a_i e il secondo addendo è il valore $Q_e(a_i)$ che contiene $N = \sum_{j=1}^{K} n_j$, il numero totale di flussi di tutte le azioni.

Ad ogni passo del processo decisionale si seleziona l'azione con il maggior Q, quindi di nuovo abbiamo una politica greedy, che sceglie sempre e solo l'azione che in quel momento ha il maggior valore stimato.

Politica UPPER CONFIDENCE BOUND
Input Per ogni azione a_i: media osservata μ_i, e numero n_i dei flussi di a_i. *Parametri* Nessun parametro.
Inizializzazione Seleziona una volta ciascuna azione a_i.

[86] Auer 2002

Iterazione

Per ogni a_i:

{

$$Q(a_i) \leftarrow \mu_i + \sqrt{\frac{2 \log N}{n_i}}$$

$$a^* = \arg\max_a Q(a)$$

$$\pi(a^*) = 1$$

$$\pi(a) = 0 \; per \; a \neq a^*$$

}

Nella politica UPPER CONFIDENCE BOUND il valore di esplorazione è maggiore per le azioni meno esplorate, come in INTERVAL ESTIMATION, ma non dipende dalla dispersione dei loro flussi. La forma del valore di esplorazione non è affatto intuitiva; discende da una complessa dimostrazione matematica che ne assicura importanti proprietà statistiche.

Al crescere del numero totale di decisioni *N*, la politica converge sulla azione ottimale, selezionandola quasi sempre, salvo sempre più sporadiche esplorazioni delle altre. Abbiamo definito in precedenza il rimpianto come la differenza fra il valore totale dei flussi ottenuto in concreto da una politica e il valore ideale raggiungibile da una politica “oracolare” che sempre, fin dal primo passo, segue solo l’azione ottima. La politica appena descritta garantisce, al tendere di *N* all’infinito, che il rimpianto tenderà a crescere solo in modo logaritmico rispetto a *N*, quindi molto lentamente. In sintesi, questa politica è asintoticamente ottimale.

Questi sono risultati matematici notevoli. Resta da vedere il comportamento pratico di questa politica. Infatti non basta che la politica converga verso la politica oracolare all’infinito, occorre che lo faccia con una certa rapidità. Anzi, dopotutto una politica che converge velocemente a una performance buona può essere preferibile a una che converge lentamente a una performance ancora migliore. I processi decisionali reali in genere non sono di durata infinita.

I risultati sperimentali riportati in letteratura indicano che la convergenza di questa politica può essere troppo lenta in molte applicazioni, perché lungo il cammino tende ad essere troppo esplorativa e a convergere con troppa prudenza sull'azione greedy. Si tratta di valutazioni da prendere con prudenza, perché le sperimentazioni riportate in letteratura sono sempre su dati sintetici. L'esperienza riportata nel caso di studio in questo lavoro tende a confermare quanto appena detto, ma anche in questo caso si tratta pur sempre di evidenza aneddotica. In ogni caso, la questione non è troppo importante in pratica, perché non risulta che questo metodo sia usato in sistemi reali, anche per motivi diversi dalla velocità di convergenza. L'idea è però servita per sviluppi interessanti, che andiamo a esaminare.

9.11 Politica TUNED UPPER CONFIDENCE BOUND

La versione base della politica UPPER CONFIDENCE BOUND stima il valore di esplorazione *Qe* tenendo conto solo di una fonte di incertezza riguardo a una azione, l'ampiezza del campione di flussi di cui si usa la media, ma non la dispersione dei valori dei flussi. Sono stati quindi definiti molti metodi che includono la varianza nella formula e per questo motivo sono definiti *tuned* nel senso che sono regolati azione per azione.

Per illustrare il principio in modo semplice, supponiamo che i flussi siano bernoulliani, cioè che possano assumere i valori 1 o 0. In questo caso la media empirica dei flussi μ_i di una azione è uno stimatore della probabilità che il flusso valga 1, mentre la varianza è $\sigma_i^2 = \mu_i(1-\mu_i)$. Il valore di una azione è

$$Q(a_i) = \mu_i + \sigma_i \sqrt{\frac{\log N}{n_i}}$$

La nuova formula favorisce le azioni che sono meno campionate o che hanno maggiore varianza. Per la precisione, il valore della varianza non deve essere nullo, quindi in realtà una varianza troppo piccola viene sostituita da un limite inferiore, per esempio

$$Q(a_i) = \mu_i + min\left\{\sigma_i, \frac{1}{100}\right\} \cdot \sqrt{\frac{\log N}{n_i}}$$

Nel caso di flussi non bernoulliani la formula si complica, mantenendo però lo schema di base in cui la deviazione standard moltiplica la quantità sotto radice che decresce all'aumentare dell'ampiezza del campione.

Formule di questo tipo non permettono in generale di dare garanzie matematiche precise sulla convergenza della politica, ma abbiamo già detto che in pratica questo non è sentito come un problema troppo stringente, mentre riceve molta più attenzione la velocità con cui la politica raggiunge una ragionevole performance.

9.12 Limiti delle politiche non adattive

Le politiche viste finora, e molte altre che non è il caso di esaminare in questa sede, sono state concepite e studiate in ambito di ricerca scientifica, e purtroppo dal punto di vista del progettista di un sistema decisionale soffrono di seri limiti dovuti a questa loro origine.

Le sperimentazioni sono su dati simulati, e sistematicamente dimostrano la superiorità di una nuova politica, salvo essere in seguito contraddetti da nuovi risultati di simulazioni. Decisamente si avverte la carenza di sperimentazioni su dati reali significativi e rappresentativi di problemi diffusi.

Le ipotesi su cui si fondano sia le simulazioni sia i teoremi di convergenza e limitazione del rimpianto sono matematicamente eleganti, ma non risultano molto realistiche e quindi limitano la loro applicabilità. Fra queste ipotesi ne mettiamo in evidenza due:

1. *Stazionarietà*. Alcune politiche gestiscono la non stazionarietà meglio di altre, ma in generale sono concepite per processi stazionari. La ricerca sull'applicazione di questi metodi a processi non stazionari ha prodotto risultati scientificamente stimolanti, ma che al momento non risultano essersi diffusi nella pratica industriale.
2. *Azioni prefissate*. Si suppone che durante tutto il processo le azioni selezionabili siano sempre le stesse. In processi reali anche questa ipotesi è spesso non realistica. Anche in questo caso, la ricerca (peraltro ridotta) non sembra avere prodotto risultati applicati in sistemi reali.

L'applicazione principe per il MAB è per comune sentire la pubblicità online, ma proprio in questo settore le due ipotesi vengono meno in modo inequivocabile. Nel caso più naturale, ogni azione corrisponde a un annuncio. Gli annunci hanno un ciclo di vita con flussi decrescenti nel tempo, e continuamente nascono nuovi annunci e alcuni esistenti muoiono. In queste condizioni le politiche viste finora appaiono severamente limitate per loro natura.

Ciò non significa affatto che la comprensione di queste politiche sia priva di importanza pratica. Al contrario, esse forniscono al progettista di sistemi decisionali un quadro concettuale prezioso, che però non può essere immediatamente applicato ai problemi concreti, almeno non a quelli di rilevante valore economico, come appunto la pubblicità online.

Aggiungiamo ora un ulteriore problema di queste politiche: i loro parametri sono ben difficili da regolare manualmente, l'aggiustamento da parte di un decisore umano è troppo complesso e troppo lento. Trovare una soluzione a questo problema è, a nostro parere il punto critico per l'uso pratico dello schema MAB. La nostra tesi è che le politiche in sé sono efficaci e utilissime *purché di momento in momento i loro parametri siano adattati automaticamente*.

9.13 Esplorazione a molti stati

Le politiche sinora esaminate sono relative al problema a un solo stato. Nel nostro modello decisionale la presenza di più stati non è necessaria, e alcuni importanti problemi di ottimizzazione possono essere affrontati nella pratica prescindendo dallo stato. Ciò non toglie che in un processo decisionale basato sugli stati il dilemma dell'esplorazione resti di importanza centrale.

Un modo immediato e naturale di risolvere il dilemma dell'esplorazione in molti stati è di risolverlo separatamente in ogni stato. I dati storici sono etichettati in base allo stato del momento in cui si è scelta una azione e si è generato il relativo flusso di reddito. In questo modo il database storico è suddiviso in tanti database quanti sono gli stati, e su ciascuno si procede con una delle politiche viste. In tal modo per ogni stato si ha una procedura per scegliere una azione tenendo conto dle problema della esplorazione.

Questo modo di procedere è evidentemente corretto, ma è anche chiaro che in questo modo l'informazione non viene gestita nel modo più efficace ed efficiente. Ciò che si impara su uno stato non è utile per prendere decisioni migliori in un altro stato, e lo stesso vale per azioni simili o anche identiche selezionate in stati diversi. Possiamo dire che non c'è generalizzazione dell'informazione.

Questo problema può diventare molto grave quando gli stati sono molti e i dati osservati non sufficienti a valutare i singoli stati con campioni sufficienti.

Nel seguito del nostro lavoro proporremo alcuni approcci a questo problema.

10 METODI AVANZATI DI ESPLORAZIONE

In questo capitolo avanzeremo due proposte originali per affrontare il dilemma dell'esplorazione in processi decisionali complessi, quando le politiche descritte nel capitolo precedente non appaiono del tutto soddisfacenti.

10.1 Meta-politiche e autoregolazione dei parametri

Abbiamo visto come le politiche di esplorazione non adattive vanno incontro a seri problemi e limitazioni in ambienti non stazionari. In questa sezione proponiamo un nuovo modello di controllo per il problema dei banditi, basato sul concetto di *meta-livello*. L'idea si può esprimere così: una politica controlla la selezione delle azioni utilizzando parametri scelti da una meta-politica, che vede tali parametri come azioni.

Un esempio darà concretezza a questo enunciato.

Prendiamo in considerazione la politica ε-GREEDY. La convinzione diffusa nella comunità scientifica è che tale politica, nella sua semplicità, sarebbe in grado di soddisfare le esigenze pratiche se solo si riuscisse a scegliere di momento in momento il "giusto" parametro ε. Ma il problema è proprio che in pratica non si riesce a ottimizzare questa scelta, e in definitiva questa politica è raramente vicina all'ottimalità, benché spesso riesca a dare risultati ragionevoli per gli scopi pratici. Non si tratta solo di trovare un buon valore per il parametro, ma di adattarlo continuamente all'evoluzione del sistema da controllare.

La nostra proposta è che il parametro sia scelto da una seconda politica, una meta-politica, che può essere una qualsiasi di quella viste, anche una seconda ε-GREEDY.

Nella pratica i tipici valori per il parametro ε variano fra 0,05 e 0,25: fra il 5% e il 25% delle azioni sono scelte non perché ritenute migliori, ma perché si vuole acquisire nuova conoscenza. Non c'è nessuna ragione teorica per queste scelte, e se il sistema si dimostra relativamente stazionario allora è logico diminuire il valore di ε fino quasi a *0*. Ma allora dobbiamo stabilire che cosa significa "quasi stazionario" e come la misura del grado di stazionarietà si traduce in una scelta di un nuovo valore di ε, compiti assai complessi. D'altra parte l'importanza della questione è evidente. Se le azioni consistono in proposte di annunci pubblicitari online, porre il parametro al valore del 10% significa che nel 10% dei casi la performance degli annunci peggiorerà di una percentuale pari alla differenza fra la performance del leader e quella media. Poniamo che l'annuncio leader, quello con il migliore tasso di click storicamente osservato, abbia performance *100* e che il tasso di click medio di tutti gli annunci sia *50* (si tratta di una proporzione realistica). Se abbiamo ε = 10%, allora la politica ε-GREEDY avrà una performance

$$0{,}90 \times 100 + 0{,}10 \times 50 = 95$$

Nel 90% dei casi si sceglie l'annuncio leader e la performance è 100; nel 10% dei casi si sceglie un annuncio a caso e la performance è 50; la performance finale è 95. Quindi una buona o cattiva scelta del parametro influisce per alcuni punti percentuali sulla performance degli annunci.

Ma l'azienda che ospita gli annunci sul suo sito viene remunerata proprio in proporzione a questa performance, quindi il suo profitto varia di alcuni punti percentuali solo per la scelta non felice del parametro dell'algoritmo. In realtà la situazione è ancora più a rischio di quanto dica questo esempio, perché se la situazione non è statica e le performance degli annunci variano con il tempo, allora l'impatto della scelta del parametro può essere ben più forte di questo, che già di per sé è significativo.

Immaginiamo allora di avere una meta-politica ε-GREEDY che regola la scelta fra due meta-azioni: la prima e_1 consiste nello scegliere il valore *0* come parametro della politica di base, la seconda e_2 il valore *0,25*. Ogni volta che la politica di base seleziona un annuncio *a*, si registra sul database che tale annuncio ha ricevuto una nuova *impression* e anche un nuovo click, se questo accade:

$$impression_{t+1}(a) \leftarrow impression_t(a) + 1$$

$$r_t = 1 \; se \; cliccato, 0 \; altrimenti$$

$$click_{t+1}(a) \leftarrow click_t(a) + r_t$$

$$Q_{t+1}(a) = \frac{impression_{t+1}(a)}{click_{t+1}(a)}$$

Fin qui, è solo il normale aggiornamento del valore di una azione. In parallelo, la meta-politica aggiorna il valore della propria meta-azione, che non è la scelta di un annuncio, ma del parametro che regola la scelta dell'annuncio. Se la scelta dell'annuncio è stata fatta in un momento in cui il parametro scelto era $\varepsilon = 25\%$, vale a dire in un momento in cui era valida la scelta e_2 allora la meta-politica aggiorna il proprio database in questo modo:

$$impression_{t+1}(e_2) \leftarrow impression_t(e_2) + 1$$

$$r_t = 1 \; se \; cliccato, 0 \; altrimenti$$

$$click_{t+1}(e_2) \leftarrow click_t(e_2 a) + r_t$$

$$Q_{t+1}(e_2) = \frac{impression_{t+1}(e_2)}{click_{t+1}(e_2)}$$

Le due meta-azioni relative ai parametri sono trattate esattamente come le azioni originarie relative agli annunci. La meta-politica sceglie fra loro usando un meta-parametro *meta-ε* che è esattamente un altro ε; può quindi sembrare che il problema sia solo stato spostato da una politica all'altra, ma non è così, perché il meta-parametro ha effetti indiretti sulla selezione delle azioni. Possiamo dire che il rischio di scegliere un valore non buono del parametro di esplorazione è distribuito su più parametri: i due ε e il *meta*-ε che aiuta a sceglierli. Ci sono quindi più probabilità di avvicinarsi al valore migliore.

In una fase di quasi-stazionarietà il meta-parametro *0* avrà performance migliore del suo rivale *25%*, verrà scelto più spesso e il risultato sarà che nella politica di base il parametro ε medio tenderà verso lo *0*. In una fase dinamica il meta-valore *25%* sarà utilizzato più spesso e la politica di base tenderà ad avere un ε più grande, e quindi sarà più esplorativa, come è giusto. In generale, avere due livelli di parametri permette di "assorbire" gli errori di un livello, che sono corretti dall'altro.

Il progettista di un sistema a due livelli deciderà quali saranno le meta-azioni e i meta-parametri. Potrà ad esempio scegliere più di due valori per *meta*-ε, e così facendo otterrà una auto-regolazione più fine dei parametri. Queste scelte, benché non siano tecnicamente più semplici di quelle normali per una architettura a unico livello, sono in realtà più facili perché meno rischiose, appunto perché ci sono più meccanismi di correzione.

Il meta-livello può essere implementato con una politica diversa dal livello base: un livello ε-GREEDY può essere controllato da un livello UPPER CONFIDENCE BOUND o viceversa.

10.2 Simulazione sperimentale

L'idea del controllo a meta-livello è stata sperimentata su dati simulati. Qui vengono riportati in sintesi i risultati. In Appendice si troveranno maggiori dettagli.

Si sono costruiti 29 scenari di simulazione. In ogni scenario ci sono 10 annunci. Il tipo di problema varia in ogni scenario. Alcuni scenari sono più semplici da affrontare per un software di ottimizzazione, altri meno. I 29 scenari sono costruiti in modo da simulare situazioni relativamente realistiche.

Vengono effettuate un milione di selezioni di annuncio per ogni scenario. L'esperimento è ripetuto per 100 volte.

Le tabelle riporta il numero medio di click ottenuto da varie politiche su tutti gli scenari in tutte le simulazioni.

La prima tabella è relativa alla politica ε-GREEDY con 7 diversi valori del parametro.

Politica	**Parametro**	**Click**
ε-Greedy	ε = 0,01	4036
ε-Greedy	ε = 0,05	4192
ε-Greedy	ε = 0,10	4184
ε-Greedy	ε = 0,15	4138
ε-Greedy	ε = 0,20	4069
ε-Greedy	ε = 0,25	3999
ε-Greedy	ε = 0,30	3928

La politica ε-GREEDY ottiene i risultati migliori con parametro del 5% o 10%. Il valore 1% non dà sufficiente peso all'esplorazione, mentre i valori più alti trovano sì l'azione ottimale, ma spendono risorse eccessive in seguito con esplorazione superflua. Questi valori dipendono dalla particolare scelta di dati simulati e non hanno alcuna validità generale.

La tabella successiva riporta la performance di una architettura a due livelli, in cui il livello meta è a sua volta una politica ε-GREEDY. Le azioni del livello meta sono 7, ciascuna che assegna al parametro del livello base uno dei 7 valori visti. Si sono provati 3 valori del meta-parametro ε.

Politica	**Parametro (meta)**	**Click**
ε-Greedy + ε-Greedy	ε = 0,05	4364
ε-Greedy + ε-Greedy	ε = 0,15	4372
ε-Greedy + ε-Greedy	ε = 0,25	4372

I 3 risultati sono pressoché identici e sono superiori a quelli dell'architettura a un solo livello. L'indicazione è che l'introduzione del secondo livello apporta sia benefici in performance, sia in semplicità, poiché la scelta del meta-parametro è ininfluente.

Si è poi sperimentata una architettura a due livelli in cui il secondo livello è una politica UCB (Upper Confidence Bound). Le meta-azioni sono ancora corrispondenti ai 7 possibili valori del parametro ε al primo livello. Il meta-livello UCB è stato provato con 3 diversi valori del suo parametro di esplorazione.

Politica	**Parametro (meta)**	**Click**
ε-Greedy + UCB	α = 0,02	4367

ε-Greedy + UCB	$\alpha = 0{,}05$	4377
ε-Greedy + UCB	$\alpha = 0{,}08$	4381

I risultati sono lievemente migliori del caso precedente, ma non in modo significativo. Il parametro sembra di nuovo poco influente. L'indicazione che emerge è ancora la stessa: migliore performance e maggiore semplicità rispetto a una politica tradizionale.

L'esperimento successivo si riferisce a una architettura in cui la politica base è una UCB e la politica meta è una ε-GREEDY. Le meta-azioni sono relative ai tre possibili valori del parametro α al livello base, cioè i già visti 2%, 5%, 8%. Si sono provati 3 valori del meta-parametro ε, il 5%, il 15% e il 25%.

Politica	**Parametro (meta)**	**Click**
UCB + ε-Greedy	$\varepsilon = 0{,}05$	4394
UCB + ε-Greedy	$\varepsilon = 0{,}15$	4443
UCB + ε-Greedy	$\varepsilon = 0{,}25$	4419

I risultati sono migliori; il meta-parametro ha qualche influenza, m ancora non decisiva.

Infine l'architettura a due livelli UCB, con i valori dei parametri già visti.

Politica	**Parametro (meta)**	**Click**
UCB + UCB	$\alpha = 0{,}02$	4468
UCB + UCB	$\alpha = 0{,}05$	4491

UCB + UCB	α = 0,08	4473

Nel complesso, la politica UCB sembra essere più efficace della politica ε-GREEDY a entrambi i livelli.

Questi risultati devono essere considerati come indicazioni, non prove. Gli scenari utilizzati sono ancora troppo artefatti per rappresentare la complessità di un sistema reale: gli annunci sono pochi e un milione di selezioni è un numero ridotto. Tuttavia, è lecito pensare che una maggiore complessità esalti l'effetto positivo dello schema a due livelli. Una sperimentazione più estesa può essere un interessante sviluppo di ricerca della proposta qui presentata.

10.3 Decisioni contestuali

Il problema contestuale dei banditi (CMAB) è a un livello di complessità intermedio fra il problema generale dei processi decisionali markoviani, a cui si richiama il nostro modello decisionale, così come il problema dei banditi classico.

Nel CMAB non ci sono stati nel pieno senso della parola, però al momento di prendere una decisone si osservano dei *dati di contesto* che rappresentano fenomeni che influenzano il valore delle azioni. Riprendendo la metafora delle slot-machine (i many-armed bandit), è come se ogni volta che dobbiamo scegliere la leva da tirare si accendesse prima una luce colorata: il colore ci dà un indizio sul valore di ogni leva. Può accadere che una leva che genera grandi flussi quando la luce è azzurra abbai una performance economica molto peggiore quando la luce è rossa.

Il contesto è quindi simile allo stato perché il valore atteso di una azione dipende dal contesto. La differenza è che una volta compiuta l'azione, il contesto della selezione successiva non dipende dal contesto e dall'azione presenti, mentre per gli stati questa dipendenza esiste. I contesti, quindi, non hanno memoria.

Per affrontare questi casi ci sono due soluzioni naturali: la prima è di attivare un diverso problema MAB per ogni contesto, la seconda di creare un modello predittivo per ogni contesto ricorrendo alle normali tecniche di predizione, per fare un esempio fra tanti a una regressione lineare.

Creare un MAB per ogni contesto è una operazione semplice ma “dispendiosa” in termini di informazione. Se ci sono *K* azioni e per ognuna *M* contesti, le osservazioni dei flussi generati dalle azioni vengono suddivise in $K \times M$ campioni, la cui significatività statistica può essere insufficiente, e in ogni caso si perde l’analogia fra azioni diverse per uno stesso contesto e l’analogia fra contesti diversi per una stessa azione.

Usare tecniche standard di predizione senza le politiche inventate appositamente per gestire al meglio il dilemma dell’esplorazione è a sua volta una rinuncia che si vorrebbe evitare se possibile. Occorre allora trovare qualche modo per inserire anche i contesti nel quadro concettuale dei MAB. La letteratura propone alcune soluzioni, certamente interessanti ma anche troppo complesse per le comuni applicazioni. In questa sede proponiamo un approccio risolutive che può semplificare molto il problema, quantomeno in una consistente gamma di casi.

L’idea è che si possa definire una funzione di *similarità* fra due contesti:

$$sim(C_1, C_2) \in [0,1]$$

La similarità assume valori fra 0 e 1. Il valore è tanto più vicino a 1 quanto più le stesse azioni generano flussi simili se selezionate nei due contesti. Questa definizione presuppone che i due contesti rendano possibili le stesse azioni[87].

Prendiamo ancora una volta esempio dalla pubblicità online. Varie pagine su vari siti web chiamano un server per avere un annuncio da visualizzare ai visitatori. Il server sceglie una fra molte campagne pubblicitarie e invia al sito chiamante il relativo annuncio. In questo esempio il contesto è la pagina che chiama, l’azione è la campagna selezionata. Il flusso di reddito è il flusso di risposte positive (il visitatore fa click sull’annuncio) o negative (il visitatore chiude la pagina senza fare click); in altri termini, ogni flusso di reddito vale 0 oppure 1.

---

[87] Il requisito non deve essere inteso in senso assoluto, perché si possono definire funzioni di similarità anche se le azioni selezionabili non sono esattamente le stesse nei due contesti.

Due pagine *P* e *Q* sono simili se le stesse campagne ricevono click in proporzione simile su di esse. La similarità fra le pagine può essere misurata confrontando uno a uno i tassi di click (*clickthrough rate*, *CTR*) delle stesse campagne nelle due pagine. Un metodo classico è di calcolare la *distanza* fra le due pagine come media dei quadrati delle differenze fra i CTR:

$$dist(P, Q) = \frac{1}{K}\sum_{a}[CTR(a, P) - CTR(a, Q)]^2$$

dove *K* è il numero di campagna comuni alle due pagine.

Una misura di distanza varia fra 0 e ∞; per convertirla in una misura di similarità si può applciare una trasformazione decrescente come

$$sim = 2^{-dist}$$

che varia fra 0 e 1.

Un modo diretto per calcolare la similarità può essere la misura *coseno*:

$$sim(P, Q) = \frac{\sum_a [CTR(a, P) \cdot CTR(a, Q)]}{\sum_a CTR(a, P)^2 \cdot \sum_a CTR(a, Q)^2}$$

Prescindendo dal particolare modo di calcolare la similarità, possiamo utilizzare questa misura per generalizzare l'informazione fra contesti secondo una logica del tutto compatibile con l'approccio dei MAB e del nostro modello decisionale.

Quando selezioniamo la campagna *a* per una visualizzazione di annuncio sulla pagina *P*, registriamo come sempre nel database delle osservazioni una nuova *impression* e un flusso di click che vale 0 oppure 1; questi dati servono ad aggiornare il valore di *a* in *P*. Poi, per ogni pagina *Q*, registriamo una quota di impression pari a *sim(P, Q)*, che è un numero fra 0 e 1, e un flusso pari a *sim(P, Q)·r,* dove *r* vale 1 o 0 secondo che ci sia stato o meno il click.

In questo modo ogni contesto non osservato in un certo momento viene campionato in un grado dipendente dalla sua similarità con il contesto campionato, e il suo flusso è stimato nello stesso modo. Il contesto osservato ha similarità 1 con sé stesso, e quindi viene campionato normalmente.

CONCLUSIONI

La tradizionale equazione dell'equilibrio economico, con la sua contrapposizione di costi e ricavi e la dipendenza del reddito da prezzi e quantità, racchiude il seme di sviluppi multiformi, alcuni dei quali ancora a pieno titolo argomento di ricerca non solo nel campo economico-aziendale propriamente detto, ma anche nelle discipline matematiche e informatiche.

Il Revenue Management nelle sue varie forme può essere appunto visto come disciplina che applica e risolve l'equazione di equilibrio economico in una molteplicità di situazioni di grande interesse pratico, utilizzando intensamente (ma non solo) metodi matematici e algoritmici.

In questo lavoro abbiamo esplorato una metodologia per interpretare e risolvere l'equazione alla luce dei principi del Computational Learning, e in modo particolare del Reinforcement Learning. Le motivazioni di questa ricerca sono sia teoriche sia applicative.

Da un punto di vista teorico, abbiamo dunque stabilito un collegamento fra uno degli elementi più fondamentali dell'Economia Aziendale e una disciplina come il Reinforcement Learning che ha origini culturali lontanissime, nella robotica. Le affinità concettuali emerse sono stimolanti e incoraggiano a nuove ricerche.

L'aspetto applicativo della nostra ricerca è molto evidente in applicazioni online, specie la pubblicità: qui l'importanza pratica di un approccio "robotico" alla ottimizzazione del profitto è immediatamente percepibile, e peraltro quasi inevitabile, visto che in questi casi la gestione di azienda si manifesta come sequenza di milioni di decisioni concatenate, migliaia al secondo.

Naturalmente, non è implicita alcuna pretesa che un tale approccio sia esaustivo e che escluda diverse visioni della ottimizzazione del profitto, al contrario. Anche l'ottimizzazione di campagne pubblicitarie online, con i suoi ritmi assolutamente non umani, è pur sempre da governare con criteri umani, fissando obiettivi e vincoli e suggerendo al sistema decisionale automatico delle linee strategiche. Né si intende escludere l'uso di diverse metodologie di apprendimento computazionale: il nostro modello decisionale è uno strumento in più, che arricchisce il repertorio del decisore di impresa.

Il primo contributo del nostro lavoro è dunque questo avvicinamento fra discipline culturalmente così diverse su un tema specifico, ma fondamentale.

Un altro contributo è il quadro concettuale del modello decisionale, che permette di unificare la formulazione di molti problemi specifici apparentemente distinti, che oggi sono raggruppati nel Revenue Management.

Un terzo contributo è pragmatico: la metodologia di ottimizzazione proposta è particolarmente adatta per imprese che affrontano problemi particolarmente complessi, oppure che non hanno risorse di competenze scientifiche adeguate all'uso dei tradizionali strumenti di modellistica e ottimizzazione matematica.

Infine, un contributo di tipo algoritmico, con la proposta di due metodi: uno per l'autoregolazione dei parametri della procedura di ottimizzazione e uno per la generalizzazione dell'informazione, intesa come possibilità di creare osservazioni fittizie ma verosimili per formulare previsioni basate su campioni più ampi.

BIBLIOGRAFIA

[ALPAYDIN] Alpaydin E., *Machine Learning*, MIT Press, Cambridge U.S., 2004.

[ANTOS] Antos A., Grover V., Szepesvari C., *Active learning in multi-armed bandits*, Algorithmic Learning Theory Conference, Budapest Hungary, 2008.

[AUDIBERT 2006] Audibert J. Y., Munos R. e Szepesvari C., *Use of variance estimation in the multi-armed bandit problem*, Neural Information Processing Systems Conference, 2006.

[AUDIBERT 2007] Audibert J. Y., Munos R. e Szepesvari C., *Tuning bandit algorithms in stochastic environments*, Algorithmic Learning Theory Conference, Sendai Japan, 2007.

[AUER] Auer P., Cesa-Bianchi N. e Fischer P., *Finite time analysis of the multiarmed bandit problem*, 15th International Conference on Machine Learning, 2002.

[AZZALINI] Azzalini A. e Scarpa B., *Analisi dei dati e data mining*, Springer, Milano, 2004.

[BATTAGLIA] Battaglia F., *Metodi di previsione statistica*, Springer, Milano, 2007.

[BEE DAGUM] Bee Dagum E., *Analisi delle serie storiche*, Springer, Milano, 2002.

[BERTINI] Bertini U., *Il sistema d'azienda*, Giappichelli, Torino, 1990.

[BIANCHI MARTINI] Bianchi Martini S., Cinquini L., Di Stefano G. e Galeotti M., *Introduzione alla valutazione del capitale economico*, Franco Angeli, Milano, 2002.

[BITRAN] Bitran G., Caldentey R., *An Overview of Pricing Models for Revenue Management*, Manufacturing & Service Operations Management vol. 5, no. 3.

[CAPOCCHI] Capocchi A., *La redditività aziendale – Le logiche di Revenue Management*, Franco Angeli, Milano 2008.

[CESA-BIANCHI] Cesa-Bianchi N., Lugosi G., *Prediction, learning and games*, Cambridge University Press, New York 2006.

[COQUELIN] Coquelin P. A. e Munos R., *Bandit algorithms for tree search*, Uncertainty in Artificial Intelligence Conference, Vancouver Canada 2007.

[DALLI] Dalli D., Romani S., *Il comportamento del consumatore*, Franco Angeli, Milano, 2003.

[GHIANI] Ghiani G. e Musmanno R. (a cura di), *Modelli e metodi decisionali in condizioni di incertezza e rischio*, McGraw-Hill, Milano, 2009.

[GIANNESSI] Giannessi E., *Il "Kreislauf" tra costi e prezzi*, Giuffrè, Milano, 1982.

[GRASSO] Grasso F., *Il Revenue Management alberghiero*, Hoepli, Milano, 2006.

[HAN] Han J. E Kamber M., *Data Mining*, Academic Press, San Diego, 2001.

[HARTLAND] Hartland C., Gelly S., Baskiotis N., Teytaud O. e Sebag M., *Multi-armed bandits, dynamic environments and meta-bandits*, NIPS Workshop on Online Trading of Exploration and Exploitation, 2006.

[HOOS] Hoos H. H., Stützle T., *Stochastic local search*, Morgan Kaufmann, San Francisco 2005.

[KLEINBERG] Kleinberg R., Slivkins A. e Upfal E., *Multi-Armed Bandits in Metric Spaces*, ACM Symposium on Theory of Computing, 2008.

[KREPS] Kreps D. M., *Microeconomia per manager*, Egea, Milano, 2005.

[LEGRENZI] Legrenzi P., *Psicologia e investimenti finanziari*, Il Sole 24 Ore, Milano, 2006.

[LOCANE] Locane P. D., *Revenue management*, Marco Valerio, Torino, 2009.

[MARCHI 2003] Marchi L., *I sistemi informativi aziendali*, Giuffrè, Milano, 2003.

[MARCHI 2006] Marchi L. (a cura di), *Introduzione all'economia aziendale*, Giappichelli, Torino, 2006

[MARCHI 2009] Marchi L. e Mancini D. (a cura di), *Gestione informatica dei dati aziendali*, Franco Angeli, Milano, 2009.

[MIOLO VITALI] Miolo Vitali P., *Il sistema delle decisioni aziendali*, Giappichelli, Torino, 1993.

[MITCHELL], Mitchell T. M., *Machine Learning*, McGraw-Hill, Singapore, 1997.

[ODLYZKO] Odlyzko A. M., *Privacy, economics, and price discrimination on the Internet*, Fifth International Conference on Electronic Commerce, 2003.

[PHILLIPS] Phillips R. L., *Pricing and revenue optimization*, Stanford University Press, Stanford, 2005.

[PIATTELLI PALMARINI] Piattelli Palmarini M., *Psicologia ed economia delle scelte*, Codice, Torino, 2005.

[ROMANI] Romani S. *L'analisi del comportamento del consumatore per la determinazione del prezzo di vendita di prodotti e servizi*, Franco Angeli, Milano, 2000.

[ROSSI] Rossi P. E., Allenby G. M. e McCulloch R., *Bayesian statistics and marketing*, Wiley, Chicester, 2005.

[SCHWIND] Schwind M., *Dynamic Pricing and automated resource allocation for complex information services*, Springer Verlag, Berlin, 2007.

[SIMON] Simon H. e Zatta D., *Strategie di pricing*, Hoepli, Milano, 2006.

[SUTTON] Sutton R. S. e Barto A. G., *Reinforcement Learning*, MIT Press, Cambridge U.S., 1998.

[SZITA] Szita I. e Lorincz A., *The many faces of optimism: a unifying approach*, 25th International Conference on Machine Learning, Helsinki Finland, 2008.

[TAN] Tan P., Steinbach M. e Kumar V., *Data Mining*, Pearson Education, Boston, 2006.

[TALLURI 2004] Talluri K. T. e Van Ryzin G. J. *The theory and practice of Revenue Management*, Springer, New York, 2004.

[TALLURI 2008] Talluri K. T., Van Ryzin G. J., Karaesmen I. Z. e Vulcano G. J., *Revenue Management: models and methods*, IEEE 2008 Winter Simulation Conference.

[VERCELLIS] Vercellis C., *Ottimizzazione – Teoria, metodi, applicazioni*, McGraw-Hill, Milano, 2008.

[VERMOREL] Vermorel J. e Mohr M., *Multi-Armed Bandit Algorithms and Empirical Evaluation*, European Conference on Machine Learning, 2005.

[WANG] Wang C., Kulkarni S. R. e Poor H. V., *Bandit problems with side observations*, IEEE Transactions on Automatic Control, May 2005.

[WITTEN] Witten I. H. e Frank E., *Data Mining*, Morgan Kaufmann, San Francisco 2005.

[YEOMAN] Yeoman I. e McMahon-Beattie U., *Revenue Management and pricing*, Thomson Learning, London, 2004.

[ZAPPA] Zappa G., *Il reddito d'impresa. Scritture doppie, conti e bilanci di aziende commerciali*, Giuffrè, Milano, 1950.

[ZATTA] Zatta D., *Revenue Management*, Hoepli, Milano, 2007.