

# Business Analytics

1 CAP 1 – Introduzione alle decisioni in condizioni di incertezza	1
2 CAP 2 – Elementi di complessità computazionale	18
3 CAP 3 – Metodi di decomposizione in ottimizzazione	23
4 CAP 4 – Sistemi MRP/ERP e approccio JIT	35
5 CAP 5 – Schedulazione nella produzione e nei servizi	40
6 CAP 6 – Fondamenti microeconomici del pricing e modelli di scelta discreta	46
7 CAP 7 – Il principio della programmazione dinamica	74
8 CAP 8 – Implementazione della programmazione dinamica	84
9 CAP 9 – Modelli della programmazione dinamica	92
10 CAP 9 – Modeling for Dynamic Programming	92
11 CAP 10 – Programmazione dinamica numerica per stati discreti	99
12 CAP 11 – Programmazione dinamica approssimativa e apprendimento per rinforzo per stati discreti	103
13 CAP 12 – Simulazione	108

## 1 CAP 1 – Introduzione alle decisioni in condizioni di incertezza

**Recap** Le decisioni di business si collocano spesso in contesti caratterizzati da **incertezza**, rendendo insufficiente una semplice previsione puntuale delle variabili rilevanti. Le tre componenti della **Business Analytics** chiariscono ruoli diversi ma complementari: la descrittiva analizza ciò che è accaduto, la predittiva costruisce modelli per descrivere il futuro, mentre la prescrittiva individua **azioni ottimali** tenendo conto delle conseguenze economiche.

La qualità di una previsione dipende dalla **funzione di perdita** associata all'errore. Penalità diverse conducono a previsioni ottimali diverse: il valore atteso è appropriato solo in presenza di costi simmetrici, mentre penalità asimmetriche portano naturalmente a soluzioni basate su **quantili**. Non esiste quindi una previsione universalmente ottima, ma solo scelte coerenti con il criterio decisionale adottato.

Nei problemi decisionali l'obiettivo diventa la **scelta di un vettore di decisioni** sotto incertezza. Quando i fattori di rischio non sono probabilisticamente modellabili si adotta un approccio **robusto**, basato su scenari peggiori; quando invece è disponibile una distribuzione di probabilità si ottiene un **problema di ottimizzazione stocastica**. In generale, sostituire l'incertezza con valori medi non preserva l'ottimalità della decisione.

Il modello del **newsvendor** evidenzia come la decisione ottimale derivi dal bilanciamento tra **costi di underage e overage**, mostrando che la domanda media non è un criterio corretto. Questo esempio chiarisce il legame diretto tra struttura dei costi e decisioni sotto incertezza.

I **vincoli probabilistici** consentono di imporre il rispetto dei vincoli solo con una certa probabilità, ma possono generare insiemi ammissibili non convessi e problemi difficili da trattare. Gli **alberi decisionali** permettono invece di rappresentare decisioni **adattive**, distinguendo tra nodi decisionali e nodi casuali e risolvendo il problema tramite **backward induction** sulla base del valore atteso.

Il valore dell'informazione viene quantificato attraverso **EVPI** e **VSS**, che misurano rispettivamente il beneficio teorico dell'informazione perfetta e il guadagno ottenuto risolvendo il problema stocastico rispetto a una soluzione deterministica a valore atteso.

L'introduzione ai **modelli a due stadi** chiarisce la distinzione tra decisioni **here-and-now** e decisioni **wait-and-see**, mostrando il ruolo delle **decisioni di ricorso**. Questa struttura si estende ai modelli multistadio, nei quali le decisioni dipendono dall'intera storia delle realizzazioni aleatorie e richiedono il rispetto della **non-anticipatività**.

Nei modelli multistadio emerge il problema della **generazione degli scenari** e della crescita dell'albero, che rende necessarie tecniche di campionamento, riduzione e progettazione controllata degli scenari. La qualità di una soluzione va quindi valutata anche in termini di **stabilità in-sample e out-of-sample**, verificando che le prestazioni non dipendano in modo critico dall'albero utilizzato.

Si analizzano le **proprietà di convessità** dei modelli stocastici: i vincoli probabilistici non preservano la convessità in generale, mentre nei modelli a due stadi con ricorso lineare la funzione di ricorso risulta convessa, proprietà fondamentale per lo sviluppo di metodi di soluzione basati su decomposizione.

**Introduzione** Si distinguono tre principali tipologie di *Business Analytics*

- **Business Analytics descrittiva:** ha l'obiettivo di **analizzare** e sintetizzare i **dati storici** al fine di comprendere che cosa è accaduto in passato (e.g. overbooking nel trasporto aereo).
- **Business Analytics predittiva:** mira a **stimare eventi o variabili future** sulla base di modelli (e.g. distinzione tra *forecasting*—previsione non sotto il mio controllo— e *prediction*).
- **Business Analytics prescrittiva:** si concentra sul **supporto alle decisioni**, indicando quale azione dovrebbe essere intrapresa per ottimizzare una misura (e.g. *operations management*).

## 1.1 Motivazione: perché considerare l'incertezza

**Previsioni puntuali e decisioni** Un punto di partenza naturale nello studio delle decisioni in condizioni di incertezza è la **costruzione di una previsione puntuale** di una variabile aleatoria. Sia  $X$  una v.a. reale, con distribuzione di probabilità nota. Una **previsione puntuale** consiste nella scelta di un valore  $x \in \mathbb{R}$ , che rappresenta una **stima ex ante** della realizzazione futura di  $X$ . La **qualità di una previsione** deve essere misurata in funzione del **costo associato all'errore** di previsione, che dipende dalla discrepanza tra il valore scelto  $x$  e la realizzazione effettiva di  $X$ .

**Errore quadratico medio (MSE)** Una scelta naturale per misurare il **costo dell'errore di previsione è la perdita quadratica**. In questo caso, il costo associato alla previsione puntuale  $x$  è  $(X - x)^2$ . Il *problema decisionale* consiste quindi nel **minimizzare l'errore quadratico medio**:  $\mathbb{E}[(X - x)^2]$ . Sviluppando il valore atteso, si ottiene

$$MSE(x) = \mathbb{E}[(X - x)^2] = \mathbb{E}[X^2] - 2x\mathbb{E}[X] + x^2.$$

La condizione di ottimalità del primo ordine implica che la *previsione ottima coincide con il valore atteso della variabile aleatoria*,  $x^* = \mathbb{E}[X]$ . Questo risultato mostra che **l'uso del valore atteso come previsione puntuale è ottimale solo sotto l'ipotesi di una perdita quadratica** (simmetrica).

**Perdita assoluta e penalità asimmetriche** Un criterio alternativo consiste nel misurare l'errore di previsione tramite la **deviazione assoluta**,  $\mathbb{E}[|X - x|]$ . In questo caso, la soluzione ottima è data dalla **mediana della distribuzione** di  $X$ , anziché dal valore atteso.

Nelle applicazioni economiche, errori di previsione positivi e negativi possono avere impatti diversi. È quindi naturale introdurre una **funzione di perdita asimmetrica** del tipo

$$\mathbb{E}[c_u(X - x)^+ + c_o(X - x)^-],$$

dove  $(\cdot)^+ = \max\{0, \cdot\}$  e  $(\cdot)^- = \max\{0, -\cdot\}$ , mentre  $c_u$  e  $c_o$  rappresentano rispettivamente il **costo di sottostima e di sovrastima**. La soluzione ottima è un *quantile della distribuzione di  $X$* , il cui livello dipende dal rapporto tra  $c_u$  e  $c_o$ .

Il **messaggio essenziale** è che una *previsione puntuale*, tipicamente basata sul valore atteso, non è sufficiente per prendere decisioni ottimali in condizioni di incertezza. La previsione ottimale dipende dalla funzione di perdita adottata e quindi dall'impatto economico dell'errore di previsione. In generale, *non esiste una previsione "migliore" in senso assoluto, ma solo previsioni coerenti con uno specifico criterio decisionale*.

**Modelli decisionali in ambito business** L'obiettivo non è la previsione di una variabile aleatoria, ma la **selezione di un vettore di decisioni**  $x \in X$  che ottimizzi una funzione economica in presenza di *fattori di rischio*. Se  $\xi$  denota un vettore di variabili aleatorie che influenzano il risultato economico, il problema può essere formulato come un **problema di ottimizzazione sotto incertezza**

$$\min_{x \in S} f(x, \xi).$$

Se non si dispone di informazioni probabilistiche sui fattori di rischio e si conosce solo un insieme di incertezza  $U$ , e il set di ammissibilità  $S$ , il problema assume la forma di un **worst-case robust optimization problem**

$$\min_{x \in S} \max_{\xi \in U} f(x, \xi).$$

Se invece i fattori di rischio sono modellati come variabili casuali con distribuzione nota, si ottiene un **problema di ottimizzazione stocastica**

$$\min_{x \in S} \mathbb{E}_{\mathbb{P}}[f(x, \tilde{\xi})].$$

Questa formulazione mette in evidenza che, in generale,  $\mathbb{E}_{\mathbb{P}}[f(x, \tilde{\xi})] \neq f(x, \mathbb{E}_{\mathbb{P}}[\tilde{\xi}])$ , e giustifica la necessità di *modelli decisionali che tengano esplicitamente conto dell'incertezza*. La formulazione di questo modello suggerisce un problema decisionale statico, in base al quale prendiamo una decisione  $x$  prima della realizzazione della variabile casuale  $\tilde{\xi}$ , ma non c'è modo di adattare la decisione dopo il verificarsi dell'evento casuale.

**Differenza tra rischio e incertezza** Il **rischio** è definito come la presenza di variabili aleatorie con distribuzione di probabilità nota, mentre l'**incertezza** si riferisce a situazioni in cui tali distribuzioni non sono note o non possono essere stimate con precisione.

## 1.2 Modelli decisionali statistici

**Il modello classico del newsvendor** Il **modello classico del newsvendor** rappresenta un problema di decisione sotto condizione di incertezza in cui *una decisione deve essere presa prima dell'osservazione di una variabile aleatoria*. Sia  $D$  una variabile casuale non negativa che rappresenta la **domanda**, con distribuzione di probabilità nota. Il decisore sceglie una **quantità**  $q \geq 0$ , a **prezzo**  $c$  prima di osservare la realizzazione di  $D$ . Ogni pezzo viene venduto a un **prezzo**  $p > c$  durante la finestra di vendita, e ad un **prezzo ridotto** successivamente  $p_u < c$ .

L'intuizione potrebbe suggerire di porre semplicemente  $q = \mathbb{E}[D]$  per massimizzare il valore atteso del profitto. Ma questo risulta essere sbagliato. In realtà, il risultato non sorprende se introduciamo due tipi di costo:

- Se  $q < D$ , avremo un **costo opportunità**  $m = p - c$ , cioè il margine di profitto per la parte di domanda non soddisfatta (**shortfall**).
- Se  $q > D$ , avremo un **costo dell'invenduto**  $c_u = c - p_u$ , legata alla svendita della giacenza residua (**surplus**).

**Espressione generale del profitto e riscrittura con costi di underage/overage** La **formulazione tramite funzione di perdita** può essere affiancata da una scrittura esplicita del profitto, utile per evidenziare la decomposizione in termini di *underage* e *overage*. L'espressione generale del profitto è

$$\pi(q, d) = -cq + p \min(q, d) + p_u \max(q - d, 0) = -cq + p \min(q, d) + p_u(q - d)^+.$$

Usando l'identità  $q = \min(q, d) + (q - d)^+$ , il profitto può essere riscritto come  $\pi(q, d) = c_u \min(q, d) - c_o(q - d)^+$ , dove definiamo il costo di underage  $c_u = p - c$  e il costo di overage  $c_o = c - p_u$ .

Se l'**incertezza sulla domanda** è modellata da una v.a. continua con densità  $f_D(x)$ , il **profitto atteso** è

$$\mathbb{E}[\pi(q, D)] = c_u \left( \int_0^q x f_D(x) dx + \int_q^{+\infty} q f_D(x) dx \right) - c_o \int_0^q (q - x) f_D(x) dx.$$

**Teorema – Regola di Leibniz** Consideriamo una funzione di due variabili  $g(q, x)$  e definiamo una funzione della sola  $q$  come

$$G(q) = \int_{h_1(q)}^{h_2(q)} g(q, x) dx.$$

Notiamo che anche gli estremi di integrazione sono funzioni di  $q$ . Sotto opportune ipotesi di continuità, la **regola di Leibniz** permette di scrivere

$$\frac{dG}{dq}(q) = \int_{h_1(q)}^{h_2(q)} \frac{\partial g}{\partial q}(q, x) dx + g(q, h_2(q)) h'_2(q) - g(q, h_1(q)) h'_1(q).$$

Nel caso del **profitto atteso del newsvendor**, tramite la regola di Leibniz otteniamo

$$\begin{aligned} \frac{d\mathbb{E}[\pi(q, D)]}{dq} &= c_u \left( q f_D(q) + \int_q^{+\infty} f_D(x) dx - q f_D(q) \right) - c_o \int_0^q f_D(x) dx \\ &= c_u \int_q^{+\infty} f_D(x) dx - c_o \int_0^q f_D(x) dx = c_u(1 - F_D(q)) - c_o F_D(q) = 0, \end{aligned}$$

dove  $F_D(x) = \mathbb{P}\{D \leq x\}$  è la *funzione di distribuzione cumulativa* della domanda. Ponendo a zero la derivata prima, ricaviamo immediatamente

$$F_D(q^*) = \frac{c_u}{c_u + c_o}.$$

Come verifica, risulta utile controllare la **derivata seconda**

$$\frac{d^2\mathbb{E}[\pi(q, D)]}{dq^2} = -c_u f_D(q) - c_o f_D(q) < 0 \quad \forall q,$$

in quanto la funzione di densità non può assumere valori negativi. Ciò dimostra la **concavità del profitto atteso rispetto a  $q$** .

**Modelli con vincoli probabilistici (chance-constrained models)** L'idea alla base dei *chance-constrained models* è che un **vincolo stocastico non debba necessariamente essere soddisfatto in ogni realizzazione dell'incertezza**, ma solo con una probabilità sufficientemente elevata. In particolare, un vincolo del tipo  $g(x, \xi) \leq 0$  è considerato *accettabile* se risulta soddisfatto con probabilità almeno pari a un livello prefissato. Sono considerati **vincoli probabilistici individuali o congiunto**

$$\mathbb{P}\{g_j(x, \tilde{\xi}) \leq 0\} \geq 1 - \alpha_j, \quad j \in [m], \quad \mathbb{P}\{\mathbf{g}(\mathbf{x}, \tilde{\xi}) \leq \mathbf{0}_m\} \geq 1 - \alpha,$$

dove  $\mathbf{g}$  è una **funzione vettoriale**. L'intuizione potrebbe suggerire che, *se le funzioni  $g_j$  sono convesse rispetto a  $x$  per ogni realizzazione di  $\tilde{\xi}$ , allora anche il vincolo probabilistico dovrebbe preservare la convessità*. Tuttavia, **questo non è vero in generale**: *l'insieme ammissibile di un problema con vincoli probabilistici può risultare non convesso*, in quanto definito come unione di insiemi ammissibili associati a diversi scenari. Per questo motivo, i modelli chance-constrained possono risultare difficili da trattare dal punto di vista computazionale e spesso vengono approssimati tramite formulazioni di tipo robusto. Nonostante ciò, essi rappresentano uno strumento naturale per modellare problemi statici in cui non è possibile adattare le decisioni dopo la realizzazione dell'incertezza.

## 1.3 Alberi decisionali

**Dalle decisioni statiche a quelle adattive** Un modo naturale per introdurre **decisioni adattive in condizioni di incertezza** è l'utilizzo degli **alberi di decisione**. Gli alberi di decisione *consentono di rappresentare in modo esplicito la sequenza temporale delle decisioni e delle realizzazioni aleatorie*, evidenziando come le decisioni possano essere adattate sulla base delle informazioni che si rendono disponibili nel tempo. Un albero di decisione è costituito da due tipi fondamentali di nodi

- **Nodi decisionali**, rappresentati da quadrati, che corrispondono a **scelte discrete tra alternative mutuamente esclusive**. In questi nodi il decisore deve selezionare una sola azione tra quelle disponibili.
- **Nodi casuali**, rappresentati da cerchi, che descrivono la **realizzazione di esiti aleatori**. A ciascun esito  $i$  è associata una probabilità  $\pi_i$ , e la somma delle probabilità degli esiti che partono da uno stesso nodo casuale è pari a 1.

Un albero di decisione è composto da una *sequenza di nodi decisionali e nodi casuali*, che in genere si **alternano**. L'albero termina con **nodi terminali**, rappresentati da punti, ai quali sono associati i **risultati finali**, tipicamente espressi in termini monetari.

**Strategia e criterio di ottimalità** Risolvere un problema rappresentato mediante un albero di decisione significa individuare una **strategia**, ovvero una *regola che specifica quale decisione assumere in ciascun nodo decisionale che può essere visitato*. Assumendo che i risultati abbiano natura monetaria, il criterio più naturale è la **massimizzazione del valore monetario atteso EMV** (*Expected Monetary Value*).

Quando un nodo decisionale è seguito da nodi casuali, ciascun nodo casuale viene etichettato con il proprio **valore monetario atteso**. Questo consente di confrontare le alternative disponibili nel nodo decisionale. La procedura di soluzione procede a **ritroso nel tempo** (*backward induction*), iniziando dai nodi terminali e risalendo progressivamente verso la radice dell'albero. *Un nodo può essere valutato solo quando tutti i suoi successori sono già stati valutati*.

**Esempio: introduzione di un nuovo prodotto** Un tipico esempio di applicazione degli alberi di decisione riguarda la scelta se **lanciare un nuovo prodotto**, vendere una licenza o acquisire **informazioni aggiuntive** tramite un'indagine di mercato. L'indagine fornisce informazioni che modificano le state di conoscenza, ma non la probabilità di successo del prodotto. La soluzione ottimale si ottiene confrontando il valore monetario atteso delle diverse strategie, tenendo conto del costo dell'informazione e delle decisioni ottimali successive ai diversi esiti osservati.

## 1.4 EVPI and VSS

**Problema statico di ottimizzazione stocastica** Si consideri il valore ottimo di un problema di ottimizzazione stocastica statica:

$$f^* = \min_{x \in S} \mathbb{E}_P [f(x, \tilde{\xi})].$$

In questo contesto la decisione viene presa **here-and-now**, prima della realizzazione dell'incertezza.

Ci si può chiedere come migliorerebbe il valore ottimo se fosse possibile **posticipare la decisione** e osservare prima lo scenario realizzato. Formalmente, ciò equivale a scambiare l'operatore di minimo con il valore atteso:

$$f_{PI}^* = \mathbb{E}_P \left[ \min_{x \in S} f(x, \tilde{\xi}) \right].$$

Il pedice  $PI$  indica che questo valore corrisponde all'ottimo ottenibile in presenza di **informazione perfetta**.

**Valore atteso dell'informazione perfetta (EVPI)** Per un problema di minimizzazione vale la disegualanza  $f_{PI}^* \leq f^*$ , poiché l'accesso all'informazione perfetta non può peggiorare la soluzione. La differenza tra i due valori ottimi definisce il **valore atteso dell'informazione perfetta**:

$$EVPI = f^* - f_{PI}^*.$$

L'EVPI fornisce una misura quantitativa dell'**impatto dell'incertezza** sul problema decisionale. L'EVPI misura il beneficio teorico della **chiaroveggenza**. In pratica, l'accesso a informazione perfetta è estremamente raro; pertanto, l'EVPI va interpretato come un **limite superiore** al valore che l'informazione può avere.

**Esempio: scelta di una strategia di investimento** Un esempio illustrativo riguarda la scelta tra due strategie di investimento alternative, una *aggressiva* e una *conservativa*, in presenza di diversi stati possibili dell'economia. Nel caso **here-and-now**, la decisione viene presa prima di osservare lo stato dell'economia e la strategia viene scelta massimizzando il rendimento atteso. Nel caso **wait-and-see**, la strategia ottimale può essere selezionata dopo aver osservato lo scenario realizzato, ottenendo un valore atteso maggiore. La differenza tra i due valori rappresenta l'EVPI.

**Soluzione a valore atteso** Un approccio più semplice consiste nel **trascurare l'incertezza** e sostituire le variabili aleatorie con i loro valori attesi. In questo modo si ottiene il problema deterministico a valore atteso:

$$f_{EV}^* = \min_{x \in S} f\left(x, \mathbb{E}_P[\tilde{\xi}]\right),$$

la cui soluzione è detta **expected-value solution** e viene indicata con  $\bar{x}$ .

La soluzione  $\bar{x}$  deve essere valutata nel contesto stocastico originale. Il costo risultante è la variabile aleatoria  $f(\bar{x}, \tilde{\xi})$ , di cui si considera il valore atteso:

$$f_{EEV} = \mathbb{E}_P[f(\bar{x}, \tilde{\xi})].$$

Il confronto corretto non è tra  $f_{EV}^*$  e  $f^*$ , ma tra  $f_{EEV}$  e  $f^*$ . Infatti, il **valore della soluzione stocastica** è definito come

$$VSS = f_{EEV} - f^*,$$

nel caso di minimizzazione (con scambio dei termini nel caso di massimizzazione). È possibile dimostrare che  $VSS \geq 0$ . Quando il VSS è elevato, *lo sforzo computazionale richiesto per risolvere il problema stocastico completo è giustificato dal miglioramento ottenuto rispetto alla soluzione deterministica a valore atteso*.

**Collegamento con modelli multi-stadio** Il concetto di VSS viene ulteriormente chiarito attraverso esempi numerici e rappresenta un ponte naturale verso i **modelli di ottimizzazione a due stadi e multi-stadio**, nei quali le decisioni possono essere adattate dopo l'osservazione dell'incertezza.

## 1.5 Un'introduzione a un modello a due stadi: Assemble-to-order (ATO)

**ATO** Si consideri un esempio didattico di **pianificazione della produzione** in cui i prodotti finiti sono ottenuti assemblando un insieme di componenti. Ogni prodotto finale è caratterizzato da un insieme di caratteristiche, per ciascuna delle quali sono disponibili opzioni alternative che richiedono componenti diversi. Dalla combinazione di un numero limitato di componenti può derivare un numero molto elevato di prodotti finali.

Per questo motivo, non è possibile proteggersi dall'incertezza della domanda mantenendo scorte di prodotti finiti. D'altra parte, una strategia puramente **make-to-order** può risultare poco efficace a causa dei lunghi tempi di approvvigionamento dei componenti. In un ambiente **assemble-To-Order**, in cui l'assemblaggio è rapido, è invece possibile mantenere a magazzino i componenti e assemblare i prodotti finali solo dopo aver osservato la domanda.

Questo contesto suggerisce naturalmente una **strategia a due stadi**:

- Pianificazione della produzione dei componenti **here-and-now**, sotto incertezza sulla domanda.
- Assemblaggio dei prodotti finali in modalità **wait-and-see**, dopo l'osservazione degli ordini.

**Modello deterministico a valore atteso** *Ignorando l'incertezza della domanda e sostituendola con il valore medio*, si introducono le seguenti variabili decisionali

- $x_i \in \mathbb{Z}_+$ ,  $i = 1, \dots, n_i$ , numero di componenti prodotti.
- $y_j \in \mathbb{Z}_+$ ,  $j = 1, \dots, n_j$ , numero di prodotti finali assemblati.

Il **problema di massimizzazione del profitto** può essere formulato così

$$\begin{aligned} \max \quad & - \sum_{i \in [n_i]} C_i x_i + \sum_{j \in [n_j]} P_j y_j \\ \text{s.t.} \quad & \sum_{i \in [n_i]} T_{im} x_i \leq L_m, \quad m \in [n_m] \\ & y_j \leq \bar{d}_j, \quad j \in [n_j] \\ & \sum_{j \in [n_j]} G_{ij} y_j \leq x_i, \quad i \in [n_i] \\ & x_i, y_j \in \mathbb{Z}_+, \quad i \in [n_i], j \in [n_j]. \end{aligned}$$

**Limiti della soluzione deterministica** La soluzione ottenuta è facilmente interpretabile ma estremamente **sbilanciata**. Essa rappresenta una scommessa sulla domanda del prodotto più redditizio e può risultare molto rischiosa se la domanda effettiva si discosta dal valore medio.

**Modello stocastico a due stadi** Per tenere conto dell'incertezza, si introducono:

- Una domanda scenario-dipendente  $d_j^s$  con probabilità  $\pi_s$ ;
- Variabili di secondo stadio  $y_j^s \in \mathbb{Z}_+$  per l'assemblaggio adattivo.

Il **modello stocastico a due stadi** è

$$\begin{aligned} \max \quad & - \sum_{i \in [n_i]} C_i x_i + \sum_{s \in [n_s]} \pi^s \left( \sum_{j \in [n_j]} P_j y_j^s \right) \\ \text{s.t.} \quad & \sum_{i \in [n_i]} T_{im} x_i \leq L_m, \quad m \in [n_m] \\ & y_j^s \leq \bar{d}_j^s, \quad j \in [n_j], s \in [n_s] \\ & \sum_{j \in [n_j]} G_{ij} y_j^s \leq x_i, \quad i \in [n_i], s \in [n_s] \\ & x_i, y_j^s \in \mathbb{Z}_+, \quad i \in [n_i], j \in [n_j], s \in [n_s]. \end{aligned}$$

La soluzione stocastica produce quantità di componenti meno estreme e decisioni di assemblaggio che variano a seconda dello scenario. Le variabili di secondo stadio rappresentano piani di contingenza e non costituiscono output operativi immediati.

**Confronto corretto tra le soluzioni** Il confronto diretto tra  $f^*$  e  $f_{EV}$  non è significativo. Occorre fissare le decisioni al primo stadio e risolvere, per ciascuno scenario  $S$ , il **problema al secondo stadio**

$$\begin{aligned} R^s(\mathbf{x}^\circ) = \max & \sum_{j \in [n_j]} P_j y_j^s \\ \text{s.t.} & y_j^s \leq d_j^s, & j \in [n_j] \\ & \sum_{j \in [n_j]} G_{ij} y_j^s \leq x_i^\circ, & i \in [n_i] \\ & y_j^s \in \mathbb{Z}_+, & j \in [n_j]. \end{aligned}$$

Dove  $R^s(\mathbf{x}^\circ)$  è il **ricavo ottimo** che otteniamo nello scenario  $s$ , dato il vettore di decisione di **primo stadio**  $\mathbf{x}^\circ$ , sfruttando in modo ottimale i componenti disponibili per soddisfare la domanda. Si noti che, in questo modello, la disponibilità dei componenti  $\mathbf{x}^\circ$  è data, a seconda dei casi, dal modello **stocastico** oppure dal modello **deterministico**. In entrambi i casi, il **ricavo atteso** risultante è dato da

$$\sum_{s \in S} \pi^s R^s(\mathbf{x}^\circ).$$

## 1.6 Modelli a due stadi

**Dall'ATO ai modelli a due stadi con ricorso** Il modello *assemble-to-order* fornisce una buona introduzione ai **modelli di programmazione lineare stocastica a due stadi con ricorso** (*recourse*), che rappresentano l'esempio più semplice di modelli di ottimizzazione adattivi. In un modello a due stadi si distinguono due tipi di decisioni:

- Decisioni **here-and-now** (primo stadio)  $x$ , che devono essere prese sotto incertezza (ad esempio al tempo  $t = 0$ );
- Decisioni **wait-and-see** (secondo stadio)  $y(\tilde{\xi})$ , che possono essere prese dopo aver osservato la realizzazione dei fattori di rischio  $\tilde{\xi}$  (ad esempio al tempo  $t = 1$ ).

Le decisioni di secondo stadio sono anche dette **decisioni di ricorso**. Si osserva immediatamente che tali decisioni sono in realtà **politiche decisionali**, cioè funzioni che mappano la realizzazione dei fattori di rischio in una decisione ammissibile; pertanto appartengono a uno spazio **infinito-dimensionale**.

**Formulazione annidata del modello** Il modello può essere formulato come **due problemi di ottimizzazione annidati**

$$\begin{aligned} \min_x & c^T x + Q(x) \\ \text{s.t.} & Ax = b \\ & x \geq 0, \end{aligned}$$

dove la **funzione di ricorso** è definita come  $Q(x) = \mathbb{E}_{\tilde{\xi}}[Q(x, \tilde{\xi})]$ . Il **problema al secondo stadio** è

$$\begin{aligned} Q(x, \xi) = \min_y & q(\xi)^T y \\ \text{s.t.} & Wy = h(\xi) - T(\xi)x \\ & y \geq 0. \end{aligned}$$

**Ricorso fisso e non-linearità indotta dal ricorso** Nel problema di secondo stadio sia la decisione di primo stadio  $x$  sia la realizzazione  $\xi$  dei fattori di rischio sono dati. Risolvendo il secondo stadio per ogni realizzazione  $\xi$ , si definisce implicitamente una funzione  $y(\tilde{\xi})$ , mostrando che le variabili di secondo stadio sono effettivamente funzioni. In questa formulazione la **matrice di ricorso**  $W$  non dipende da variabili

aleatorie e si parla di **ricorso fisso** (*fixed recourse*). La formulazione evidenzia che **la programmazione lineare stocastica con ricorso è, in generale, un problema di programmazione non lineare** (poiché  $Q(x)$  dipende da  $x$  tramite un problema di ottimizzazione interno).

La funzione  $Q(x)$  è un valore atteso rispetto alla distribuzione congiunta di  $\xi$ ; se le variabili aleatorie sono continue, essa è un integrale multidimensionale. Inoltre, tale integrale riguarda una funzione che non è nota in forma chiusa, poiché è definita implicitamente dalla soluzione di un problema di ottimizzazione. In molti casi di interesse pratico si possono tuttavia dimostrare proprietà utili della funzione di ricorso, in particolare la **convessità**.

Poiché le variabili di secondo stadio sono funzioni in uno spazio infinito-dimensionale, una strategia comune consiste nella discretizzazione tramite campionamento. Si genera un **albero di scenario** (*fan*) in cui l'evento  $\omega_s$  corrisponde alla realizzazione dello scenario  $s \in S$ , con  $S$  insieme degli scenari. Se gli scenari sono **campionati via Monte Carlo semplice**, le probabilità sono uniformi:  $\pi_s = 1/|S|$  (sono possibili anche metodi più sofisticati di generazione scenari).

**Modello deterministico equivalente (LP a grande scala)** La discretizzazione conduce a

$$\begin{aligned} & \min c^T x + \sum_{s \in S} \pi_s (q^s)^T y^s \\ \text{s.t. } & Ax = b \\ & Wy_s + T_s x = h_s, \quad s \in S \\ & x, y_s \geq 0. \end{aligned}$$

**Fattibilità del secondo stadio e concetti di ricorso completo** Una questione centrale è stabilire se **il secondo stadio sia fattibile per ogni scelta delle variabili di primo stadio** e per ogni realizzazione delle variabili aleatorie. Occorre restringere l'insieme delle decisioni *here-and-now* al dominio in cui il secondo stadio è fattibile (equivalentemente, la funzione di ricorso è limitata; come usuale, a un problema infeattibile si associa costo infinito). Riscrivendo i vincoli di collegamento come

$$Wy_s = h_s - T_s x, \quad s \in S,$$

il **secondo stadio è fattibile quando il termine noto può essere espresso come combinazione conica delle colonne di  $W$** . In tal caso si parla di **complete recourse**. Se il ricorso completo non vale per decisioni arbitrarie di primo stadio, ma vale per le decisioni che soddisfano i vincoli del primo stadio, si parla di **relatively complete recourse**.

Nei problemi di business è spesso possibile introdurre flessibilità tramite penalità, in modo da evitare infeattibilità al secondo stadio.

### 1.6.1 The plant location model

**Descrizione del problema e rete bipartita** Il classico **plant location model** è il più semplice problema di **network design**. Si considera una rete bipartita con un insieme  $P$  di nodi sorgente potenziali e un insieme  $D$  di nodi domanda. I nodi sorgente rappresentano impianti produttivi che possono essere aperti oppure no; i nodi domanda possono rappresentare magazzini regionali o centri retail. In ciascun nodo domanda si realizza una domanda aleatoria  $D_j(\omega)$ ,  $j \in D$ , che va soddisfatta al costo minimo. Sono necessari i seguenti dati

- Per ogni  $i \in P$ , un **costo fisso di apertura**  $f_i$  e una **capacità**  $u_i$ .
- Per ogni  $j \in D$  e **scenario**  $s \in S$ , una domanda  $d_j^s$  (discretizzazione della domanda aleatoria).
- Per ogni arco  $(i, j)$ , un **costo unitario di trasporto**  $c_{ij}$  (costi variabili lineari).

Il **problema è naturalmente a due stadi**: le *decisioni di apertura impianti* devono essere prese sotto incertezza, mentre le *decisioni di trasporto* possono essere rimandate a dopo l'osservazione della domanda.

Si introducono le variabili

$$y_i = \begin{cases} 1 & \text{se il nodo sorgente } i \in P \text{ è aperto} \\ 0 & \text{altrimenti} \end{cases} \quad \text{e} \quad x_{ij}^s \geq 0,$$

dove  $x_{ij}^s$  è il flusso da  $i \in P$  a  $j \in D$  nello scenario  $s \in S$ . Le variabili di localizzazione  $y_i$  sono **di primo stadio** (design), mentre i flussi  $x_{ij}^s$  sono **di secondo stadio** (controllo).

Si ottiene il seguente **modello MILP stocastico a due stadi con ricorso**

$$\begin{aligned} \min \quad & \sum_{i \in P} f_i y_i + \sum_{s \in S} \pi^s \left( \sum_{i \in P} \sum_{j \in D} c_{ij} x_{ij}^s \right) \\ \text{s.t.} \quad & \sum_{i \in P} x_{ij}^s = d_j^s, \quad \forall s \in S, \forall j \in D \\ & \sum_{j \in D} x_{ij}^s \leq R_i y_i, \quad \forall s \in S, \forall i \in P \\ & x_{ij}^s \geq 0, y_i \in \{0, 1\}. \end{aligned}$$

**Scenari estremi e necessità di formulazioni elastiche** È opportuno chiedersi *cosa accade se viene incluso uno scenario estremo con domanda molto elevata*. Potrebbe non essere nemmeno possibile soddisfare tale domanda; inoltre, la pianificazione della capacità verrebbe guidata da scenari estremi ma molto improbabili, producendo soluzioni eccessivamente costose. In questi casi conviene ricorrere a formulazioni elastiche, permettendo **violazioni penalizzate dei vincoli**.

**Modello elastico con domanda non soddisfatta penalizzata** Sia  $z_j^s \geq 0$  la **quantità di domanda non soddisfatta** nel nodo  $j$  nello scenario  $S$ . Tali variabili entrano nell'obiettivo moltiplicate per un coefficiente di penalità  $\beta_j$ . Il modello diventa

$$\begin{aligned} \min \quad & \sum_{i \in P} f_i y_i + \sum_{s \in S} \pi^s \left( \sum_{i \in P} \sum_{j \in D} c_{ij} x_{ij}^s \right) + \sum_{s \in S} \pi^s \left( \sum_{j \in D} \beta_j z_j^s \right) \\ \text{s.t.} \quad & \sum_{i \in P} x_{ij}^s + z_j^s = d_j^s, \quad \forall s \in S, \forall j \in D \\ & \sum_{j \in D} x_{ij}^s \leq R_i y_i, \quad \forall s \in S, \forall i \in P \\ & x_{ij}^s \geq 0, z_j^s \geq 0, y_i \in \{0, 1\}. \end{aligned}$$

La **quantificazione delle penalità**  $\beta_j$  dipende dal significato di  $z_j^s$ : può rappresentare *domanda effettivamente non servita* (con penalità differenziate per priorità di mercato) oppure *domanda soddisfatta tramite fornitori terzi* (con penalità pari al costo effettivo di tale opzione).

**Osservazioni di modellazione** Il modello considerato è solo un primo passo verso formulazioni più realistiche (costi di trasporto non lineari, più item). Inoltre si trascura la dimensione temporale e l'uso di scorte per gestire variabilità parzialmente prevedibile della domanda, che porterebbero a un modello multistadio più impegnativo. L'output rilevante del modello rimane la **scelta del network design**.

## 1.6.2 Modello del newsvendor a due stadi

**Motivazione e struttura informativa** Si consideri una **versione a due stadi e multi-item del newsvendor problem**, soggetta a vincoli di capacità. La struttura informativa può essere rappresentata

da un albero a tre livelli: al nodo radice si decide il primo lotto; nei nodi intermedi si decide un secondo lotto sotto incertezza ridotta; nei nodi foglia si osservano le vendite effettive e si contabilizzano costi di overage/underage. Ogni nodo foglia  $n \in N_2$  ha esattamente un predecessore (antecedente)  $\alpha(n) \in N_1$ . Il nodo radice 0 è l'antecedente diretto di ogni nodo in  $N_1$ . Le decisioni dell'ultimo livello sono in realtà variabili fittizie di contabilizzazione (surplus/shortfall penalizzati), non variabili di controllo o design. Sono necessari i seguenti insiemi e parametri

- $I$ : insieme degli item (SKU).
- $K_1$  e  $K_2$ : capacità massime (numero massimo di item producibili) nel primo e nel secondo run produttivo.
- $L_i, i \in I$ : lotti minimi. Le variabili di lotto sono semicontinue: si può non produrre un item, ma se lo si produce esiste un minimo economicamente giustificato.
- Domanda  $d_i^n$  per ciascun  $i \in I$  e foglia  $n \in N_2$ , con probabilità associata  $\pi^n$ .
- Coefficienti di penalità di overage  $c_i^o$  e underage  $c_i^u$  per ciascun  $i \in I$ .

Si introducono le **variabili decisionali**

- $x_i^n$ : quantità prodotta dell'item  $i$  nel nodo  $n \in \{0\} \cup N_1$ , intera non negativa.
- Variabili binarie  $z_i^n \in \{0, 1\}$  per rappresentare la semicontinuità.
- Variabili di contabilizzazione  $u_i^n$  (underage) e  $v_i^n$  (overage) ai nodi foglia  $n \in N_2$ .

La formulazione **nel MILP stocastico** è

$$\begin{aligned} \min \quad & \sum_{n \in N_2} \pi^n \left[ \sum_{i \in I} (c_i^o o_i^n + c_i^u u_i^n) \right] \\ \text{s.t.} \quad & \sum_{i \in I} x_i^0 \leq K_1 \\ & x_i^0 \geq m_i \delta_i^0, \quad x_i^0 \leq K_1 \delta_i^0 \quad i \in I \\ & \sum_{i \in I} x_i^n \leq K_2 \quad n \in N_1 \\ & x_i^n \geq m_i \delta_i^n, \quad x_i^n \leq K_2 \delta_i^n \quad i \in I, n \in N_1 \\ & x_i^0 + x_i^{a(n)} = d_i^n + o_i^n - u_i^n \quad i \in I, n \in N_2 \\ & x_i^n \in \mathbb{Z}_+, \quad \delta_i^n \in \{0, 1\} \quad i \in I, n \in \{0\} \cup N_1 \\ & u_i^n, o_i^n \geq 0 \quad i \in I, n \in N_2. \end{aligned}$$

### 1.6.3 Programmazione lineare stocastica multistadio con ricorso

**Generalizzazione ai modelli multistadio** Le formulazioni multistadio sorgono come generalizzazione dei modelli a due stadi. Concettualmente, si annidano **funzioni di ricorso corrispondenti ai diversi stadi decisionali**. In questo caso, le decisioni future possono dipendere dall'intera storia del processo stocastico dei fattori di rischio. Si introduce la notazione

$$\tilde{\xi}_{[t]} = (\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_t), \quad \tilde{\xi}_{[1]} = \tilde{\xi}_1.$$

La decisione *here-and-now* è  $x_0 \in X_0$  e si risolve

$$\min_{x_0 \in X_0} \left\{ f_0(x_0) + \mathbb{E}[Q(x_0, \tilde{\xi}_1)] \right\}.$$

Il termine  $Q(x_0, \xi_1)$  è definito ricorsivamente come

$$Q(x_0, \xi_1) = \min_{x_1 \in X_1(x_0, \xi_1)} \left\{ f_1(x_1, \xi_1) + \mathbb{E} \left[ Q(x_1, \tilde{\xi}_{[2]}) \mid \tilde{\xi}_1 = \xi_1 \right] \right\}.$$

Allo stesso modo, per  $t = 2, \dots, T$ , si ha

$$Q_t(x_{t-1}, \tilde{\xi}_{[t]}) = \min_{x_t \in X_t(x_{t-1}, \tilde{\xi}_{[t]})} \left\{ f_t(x_t, \xi_{[t]}) + \mathbb{E} \left[ Q_{t+1}(x_t, \tilde{\xi}_{[t+1]}) \mid \tilde{\xi}_{[t]} = \xi_{[t]} \right] \right\}.$$

Possiamo definire la funzione di ricorso

$$Q_{t+1}(x_t, \xi_{[t]}) \doteq \mathbb{E} \left[ Q_{t+1}(x_t, \tilde{\xi}_{[t+1]}) \mid \tilde{\xi}_{[t]} = \xi_{[t]} \right],$$

e riscrivere l'equazione precedente come

$$Q_t(x_{t-1}, \xi_{[t]}) = \min_{x_t \in X_t(x_{t-1}, \xi_{[t]})} \left\{ f_t(x_t, \xi_{[t]}) + Q_{t+1}(x_t, \xi_{[t]}) \right\}.$$

Nel problema dell'ultimo stadio, per  $t = T$ , si può semplicemente porre  $Q_T(x_T, \xi_{[T]}) \equiv 0$ .

Una formulazione alternativa, si ottiene enfatizzando che le decisioni di ricorso sono in realtà funzioni (*non anticipative*) del cammino campionario osservato fino a quel momento,  $x_t(\tilde{\xi}_{[t]})$ . Questo conduce alla formulazione funzionale

$$\begin{aligned} & \min_{x_0, x_1(\cdot), \dots, x_T(\cdot)} \mathbb{E} \left[ f_0(x_0) + f_1(x_1(\tilde{\xi}_{[1]}), \tilde{\xi}_1) + \dots + f_T(x_T(\tilde{\xi}_{[T]}), \tilde{\xi}_T) \right] \\ & \text{s.t. } x_0 \in X_0, \\ & \quad x_t(\xi_{[t]}), \xi_t \in X_t(x_{t-1}(\xi_{[t-1]}), \xi_{t-1}), \quad t = 1, \dots, T. \end{aligned}$$

#### 1.6.4 Un modello di pianificazione finanziaria multistadio con costi di transazione proporzionali

**Asset-liability management su albero di scenario** Si formula un modello di **asset-liability management** in cui si scambia un insieme di strumenti finanziari (asset) per far fronte a un flusso di passività (liabilities). L'incertezza è rappresentata mediante un albero di scenario.

- $L_n$  è la liability da soddisfare nel nodo  $n \in N$ .
- Non si considerano nuovi apporti di cassa lungo il percorso; l'unico modo per generare cassa è vendere asset.
- Gli asset sono indicizzati da  $i \in I$  e  $P_i^n$  è il prezzo dell'asset  $i$  nel nodo  $n$ .
- Si considerano costi di transazione proporzionali (lineari): per acquistare o vendere si paga una percentuale  $c$  del valore scambiato, sia in acquisto sia in vendita.

La radice è il nodo  $n_0$ , in cui si effettua la prima allocazione considerando le dotazioni iniziali  $\bar{h}_i^{n_0}$ . L'insieme delle foglie è  $S$ ; i nodi di trading intermedi sono  $T = N \setminus (\{n_0\} \cup S)$ . L'obiettivo è massimizzare l'utilità attesa della ricchezza terminale. Si introducono le variabili decisionali:

- $z_i^n$ : quantità di asset  $i$  acquistata nel nodo  $n$ .
- $y_i^n$ : quantità di asset  $i$  venduta nel nodo  $n$ .
- $x_i^n$ : quantità detenuta di asset  $i$  nel nodo  $n$  dopo il ribilanciamento.
- $W^s$ : ricchezza terminale nella foglia  $s \in S$ .

Le variabili  $z_i^n$ ,  $y_i^n$ ,  $x_i^n$  sono definite per  $n \in N \setminus S$  (nessun ribilanciamento alle foglie). Sia  $U(\cdot)$  l'utilità non lineare della ricchezza.

$$\begin{aligned}
\max \quad & \sum_{s \in S} \pi^s u(W^s) \\
\text{s.t.} \quad & x_i^{n_0} = \bar{h}_i^{n_0} + z_i^{n_0} - y_i^{n_0}, \quad i \in \mathcal{I} \\
& x_i^n = x_i^{a(n)} + z_i^n - y_i^n, \quad i \in \mathcal{I}, n \in \mathcal{T} \\
& (1 - c) \sum_{i \in \mathcal{I}} P_i^n y_i^n - (1 + c) \sum_{i \in \mathcal{I}} P_i^n z_i^n = L^n, \quad n \in \mathcal{N} \setminus S \\
& W^s = (1 - c) \sum_{i \in \mathcal{I}} P_i^s x_i^{a(s)} - L^s, \quad s \in S \\
& x_i^n, z_i^n, y_i^n, W^s \geq 0.
\end{aligned}$$

**Osservazioni operative** In pratica il modello verrebbe risolto ripetutamente con una logica *rolling horizon*; il ruolo dell'utilità terminale è evitare effetti distorsivi di fine orizzonte, imponendo che il portafoglio sia in una buona posizione alla fine dell'orizzonte di pianificazione. In un modello più realistico si dovrebbero includere eventuali contributi in ingresso e shortfall penalizzati rispetto alle liabilities.

## 1.7 Modelli multiperiodali e multistadio

**Esplosione dell'albero di scenario nei modelli multistadio** Un problema chiaro nei modelli di programmazione stocastica **multistadio** è la possibile **esplosione dell'albero di scenario**. Il numero di nodi necessario per rappresentare l'incertezza può diventare un fattore fortemente limitante. Questo problema è meno rilevante per una struttura ad albero **lineare**. È però importante osservare che una struttura ad albero lineare **non corrisponde a un modello multistadio**, bensì a un **modello a due stadi multiperiodale**.

**Multiperiodo non significa multistage** È possibile avere un modello multiperiodale ma **statico** se tutte le decisioni sono prese **here-and-now** e poi implementate nei periodi successivi senza adattamento al flusso informativo. Al contrario, **multistage** implica **adattamento** delle decisioni nel tempo, soggetto a vincoli di **non-anticipatività** delle politiche decisionali.

**Caso particolare del two-stage multiperiod** La struttura è peculiare: è a **due stadi**, ma coinvolge **più periodi**. Apparentemente viola la non-anticipatività, perché dopo la prima diramazione successiva alla radice sarebbe possibile prevedere perfettamente il cammino futuro. Tuttavia, una tale struttura può avere senso quando:

- Si devono prendere **decisioni strutturali o di design here-and-now** che non possono essere adattate.
- Segue un flusso di decisioni di controllo nel tempo, per le quali non si può trarre vantaggio dalla struttura lineare dell'albero.

Il modello di pianificazione energetica della sezione successiva illustra un caso di questo tipo.

### 1.7.1 Un modello multiperiodo a due stadi: Unit commitment

**Domanda aleatoria come processo discreto nel tempo** Si consideri un modello stocastico di **unit commitment**, problema di pianificazione energetica in cui l'incertezza della domanda è modellata come

processo stocastico a tempo discreto

$$\xi(\omega) = (d_1(\omega), \dots, d_T(\omega)), \quad t \in \{1, \dots, T\}, \omega \in \Omega,$$

dove  $T$  è la lunghezza dell'orizzonte di pianificazione e  $\Omega$  è l'insieme (finito) dei cammini campionari. Si denoti con  $\pi^\omega$  la probabilità dello scenario  $\omega$ .

**Classi di generatori e struttura dei costi** Si dispone di un insieme di  $I$  classi di unità generatrici. Una unità della classe  $i$  (con  $i = 1, \dots, I$ ) può produrre in un intervallo  $[m_i, M_i]$  quando è attiva. Per ciascuna classe  $i$  sono disponibili  $a_i$  unità.

Si consideri un insieme di  $I$  classi di unità di generazione, che possono fornire un livello di output nell'intervallo  $[m_i, M_i]$ ,  $i \in [I]$ , quando sono attive. Per ciascuna classe  $i$  sono disponibili  $a_i$  unità. Anche in questo caso, la potenza erogata è una variabile decisionale *semicontinua*, che può assumere valore nullo, ma che presenta un livello minimo  $m_i$  per un'unità attiva. Se un'unità di generazione della classe  $i$  è attiva al tempo  $t$ , si sostiene un costo  $E_i$  per mantenerla al livello minimo  $m_i$ , e un costo variabile lineare  $C_i$  per ogni unità addizionale di potenza erogata. Si sostiene inoltre un costo di avviamento  $s_i$  per l'accensione di un'unità della classe  $i$ .

Viene introdotta una variabile decisionale intera non negativa  $u_{it}$  che denota il numero totale di unità della classe  $i$  attive durante l'intervallo temporale  $t$ . Data la struttura dei costi, non risulta necessario rappresentare ciascuna singola unità. Si considera inoltre una variabile intera non negativa  $s_{it}$  che rappresenta il numero di unità della classe  $i$  che vengono accese all'inizio dell'intervallo temporale  $t$ . Tali variabili costituiscono decisioni di *primo stadio*. Al contrario, la variabile non negativa  $q_{it}^\omega$ , che rappresenta l'output totale della classe  $i$  durante l'intervallo temporale  $t$ , è una variabile decisionale di *secondo stadio*, indicizzata per scenario. Il **modello** può essere formulato come segue

$$\begin{aligned} \min & \sum_{i \in [I], t \in [T]} (E_i u_{it} + F_i s_{it}) + \sum_{\omega \in \Omega} \pi^\omega \left[ \sum_{i \in [I], t \in [T]} C_i (q_{it}^\omega - m_i u_{it}) \right] \\ \text{s.t.} & \sum_{i \in [I]} q_{it}^\omega \geq d_t(\omega), & t \in [T], \omega \in \Omega \\ & m_i u_{it} \leq q_{it}^\omega \leq M_i u_{it}, & i \in [I], t \in [T], \omega \in \Omega \\ & s_{it} \geq u_{it} - u_{i,t-1}, & i \in [I], t \in [T] \\ & u_{it} \leq a_i, & i \in [I], t \in [T] \\ & u_{it} \in \mathbb{Z}_+, s_{it} \in \mathbb{Z}_+, q_{it}^\omega \geq 0, & i \in [I], t \in [T], \omega \in \Omega. \end{aligned}$$

**Osservazioni critiche sul modello** Si notano due aspetti rilevanti

- **Fattibilità e scenari estremi.** Se il vincolo di domanda è imposto come vincolo rigido in ogni scenario, la soluzione può diventare molto costosa e guidata da pochi scenari improbabili con picchi di domanda elevati. Come nel plant location model, può essere opportuno introdurre una formulazione elastica, ad esempio includendo una unità “di ultima istanza” con capacità illimitata (fornitore esterno) o una penalità per domanda non soddisfatta.

- **Flusso informativo.** Il modello è multiperiodale ma a due stadi: dopo aver osservato la prima domanda al tempo  $t = 1$ , si assume di conoscere tutte le domande per  $t \in \{2, \dots, T\}$ . Questo non è realistico e si potrebbe pensare a un modello multistadio con vincoli di non-anticipatività. Tuttavia, qui non c'è modo di sfruttare tale informazione, perché le variabili di design sono *here-and-now*. Inoltre, nel modello non esiste uno stato energetico (l'energia non è immagazzinabile), e quindi i periodi sono indipendenti tra loro in termini di potenza erogata.

## 1.8 Generazione di scenari e stabilità nella programmazione stocastica

**Ruolo centrale dell'albero di scenario** Nei modelli di programmazione stocastica **multistadio**, l'elemento chiave è l'**albero di scenario**: la qualità della soluzione dipende in modo critico da quanto bene l'albero rappresenta l'incertezza. In linea di principio si potrebbe campionare un modello dinamico dei fattori di rischio tramite **simulazione Monte Carlo**, ma sorgono ulteriori difficoltà.

### Complicazioni principali

- La **dimensione del campione** deve restare contenuta, perché si associano variabili decisionali agli stati e risolvere un problema di ottimizzazione stocastica (non solo calcolare una media campionaria) può essere computazionalmente oneroso.
- Le decisioni devono essere **non-anticipative**: ciò richiede una struttura ad albero soggetta a crescita **esponenziale**. Ad esempio, con 100 rami per nodo si ottengono un milione di scenari dopo tre passi.
- Non si sta solo stimando un valore atteso, ma si stanno **prendendo decisioni** risolvendo un problema di ottimizzazione: qualunque **bias** o incoerenza nell'albero verrà sfruttata dal solver, producendo una soluzione magari buona *in-sample* ma debole *out-of-sample*.

Qualunque strategia di generazione scenari si scelga, è quindi importante controllare la **stabilità** della soluzione risultante.

**Due classi di approcci alla generazione di scenari** Si distinguono due approcci fondamentali

- **Generazione stocastica di scenari.** Basata su campionamento casuale **Monte Carlo** e quindi sulla “forza bruta” di una grande numerosità campionaria. È più gestibile nei modelli **a due stadi** che nei multistadio; le prestazioni possono essere migliorate con tecniche di **riduzione della varianza**, ad esempio *importance sampling*.
- **Generazione deterministica di scenari.** Invece della forza bruta, si selezionano scenari in modo “intelligente” tramite metodi quali **quadratura gaussiana**, **sequenze a bassa discrepanza** (ad es. **Sobol**) o procedure di generazione **ottimizzata**.

**Interpretazione come integrazione numerica (quadratura)** La generazione scenari può essere vista come un problema di **integrazione numerica**. Se  $f(x, \xi)$  dipende dalle decisioni  $x$  e da variabili aleatorie  $\xi$  con densità congiunta  $h(\xi)$ , allora

$$F(x) = \int f(x, \xi) h(\xi) d\xi \approx \frac{1}{S} \sum_{s=1}^S f(x, \xi_s).$$

In alternativa, si può usare un insieme (anche deterministico) di punti con pesi  $\{(\pi_s, \xi_s)\}$ :

$$F(x) \approx \sum_{s=1}^{S'} \pi_s f(x, \xi_s), \quad S' < S,$$

interpretando le coppie  $(\pi_s, \xi_s)$  come una **distribuzione discreta** che approssima quella continua.

**Moment/property matching e alternative** Un'idea comune è scegliere scenari e probabilità in modo da riprodurre (approssimativamente) alcune **proprietà** della distribuzione, ad esempio i **momenti** (*moment matching*). Questo approccio è stato criticato perché è possibile costruire distribuzioni diverse che condividono i primi momenti. Un'alternativa è utilizzare **metriche di distanza** tra distribuzioni: fissata una topologia dell'albero, si cercano valori e probabilità che minimizzano una distanza rispetto alla distribuzione “vera”, oppure si parte da un grande insieme di scenari e si applica **scenario reduction** per ottenere un albero più gestibile. D'altra parte, spesso ciò che conta non è la distanza tra distribuzioni, ma la **qualità della soluzione** (ad es. in un modello mean-variance può essere sufficiente approssimare bene i primi due momenti).

## Linee guida pratiche per progettare l'albero

- Se l'obiettivo è una decisione robusta di primo stadio, conviene spesso usare un branching **ricco al primo stadio** e più **limitato negli stadi successivi** (in finanza, imponendo che l'albero sia **arbitrage-free**).
- Si può ridurre la complessità usando **passi temporali non uniformi** (più fitti all'inizio, più radi in seguito).
- Si può limitare il numero di stadi **aumentando l'obiettivo** con un termine che valuta la qualità dello **terminale**, riducendo decisioni miopi e permettendo alberi più piccoli.

### 1.8.1 In- and out-of-sample stability

**Perché serve controllare la stabilità** Qualunque strategia di generazione scenari venga impiegata, è importante verificare la **stabilità** della soluzione. In pratica, l'ottimizzazione usa un albero approssimato e quindi occorre capire quanto l'output dipenda da come l'albero è stato generato.

**Problema “esatto” e problema approssimato** Il problema stocastico “esatto” può essere scritto come

$$\min_{x \in X} \mathbb{E}_{\mathbb{P}}[f(x, \xi)],$$

dove  $P$  è la **misura vera** dei fattori di rischio. In pratica si risolve un problema basato su un albero di scenario approssimato  $T$

$$\min_{x \in S} \hat{f}(x; T),$$

dove  $\hat{f}$  rappresenta una stima del valore atteso indotta dall'albero.

**Stabilità del valore obiettivo** Campionando (o costruendo) alberi diversi si ottengono soluzioni ottime diverse e valori obiettivo diversi. Spesso è più informativo controllare la stabilità del **valore dell'obiettivo** (piuttosto che della soluzione), perché quando l'obiettivo è piatto soluzioni anche molto differenti possono avere prestazioni simili.

**In-sample stability** La **stabilità in-sample** richiede che, generando due alberi  $T_1$  e  $T_2$ , i valori obiettivo ottenuti non differiscano troppo:  $\hat{f}(x_1^*; T_1) \approx \hat{f}(x_2^*; T_2)$ . Se l'albero è generato in modo deterministico, si può comunque confrontare la stabilità variando leggermente la struttura di branching per verificare che sia sufficientemente ricca.

**Out-of-sample stability** La **stabilità out-of-sample** confronta la performance reale attesa delle soluzioni sotto la distribuzione vera:  $\mathbb{E}_{\mathbb{P}}[f(x_1^*, \xi)] \approx \mathbb{E}_{\mathbb{P}}[f(x_2^*, \xi)]$ . Se gli alberi sono affidabili, le prestazioni reali non dovrebbero cambiare in modo significativo.

**Valutazione out-of-sample: due stadi vs multistadio** Nei modelli a **due stadi** la verifica out-of-sample è relativamente semplice: si fissa la soluzione di primo stadio e si risolvono molti problemi di secondo stadio corrispondenti a diverse realizzazioni di  $\xi$ . Nei modelli **multistadio**, invece, una valutazione realistica richiede tipicamente una simulazione **rolling horizon**, che è computazionalmente costosa.

### Rolling horizon e shrinking horizon

- **Sliding/rolling horizon:** si risolve un problema con  $H$  stadi, si applica la prima decisione, si osserva una realizzazione out-of-sample e si risolve di nuovo un problema con  $H$  stadi a partire dal periodo successivo.

- **Shrinking horizon**: si risolve con  $H$  stadi, si applica la prima decisione, si osserva una realizzazione e poi si risolve con  $H - 1$  stadi, riducendo progressivamente l'orizzonte. Lo shrinking horizon è più economico quando l'orizzonte di valutazione è ben definito.

**Test economico alternativo: scambio degli alberi** Un controllo più economico consiste nel generare soluzioni da due alberi  $T_1$  e  $T_2$  e **valutarle scambiando gli alberi**, verificando se  $\hat{f}(x_1^*; T_2) \approx \hat{f}(x_2^*; T_1)$ .

**Osservazione finale: limite operativo dello stochastic programming** La necessità di valutazione out-of-sample evidenzia che la programmazione stocastica produce decisioni esplicite **associate ai nodi dell'albero**; non fornisce direttamente una politica generale facilmente applicabile quando l'incertezza si realizza fuori dagli scenari previsti. Approcci come **programmazione dinamica** e **decision rules** restituiscono invece decisioni in **feedback**, più adatte alla simulazione e alla valutazione delle politiche.

## 1.9 Convessità nei modelli di programmazione stocastica

**Vincoli probabilistici congiunti** Si consideri un vincolo probabilistico congiunto della forma  $\mathbb{P}(\{\omega \mid \mathbf{g}(\mathbf{x}; \xi(\omega)) \leq \mathbf{0}\}) \geq 1 - \alpha$ , dove  $\mathbf{g}$  è una funzione vettoriale. Un punto  $\hat{\mathbf{x}}$  è ammissibile se l'insieme  $S(\hat{\mathbf{x}}) = \{\omega \mid \mathbf{g}(\hat{\mathbf{x}}; \xi(\omega)) \leq \mathbf{0}\}$  ha misura di probabilità almeno pari a  $1 - \alpha$ .

Sia  $\mathcal{F}$  il campo di tutti gli eventi e sia  $\mathcal{G} \subseteq \mathcal{F}$  la collezione degli eventi con probabilità almeno  $1 - \alpha$ . Allora  $\hat{\mathbf{x}}$  è *ammissibile* se esiste almeno un evento  $G \in \mathcal{G}$  tale che  $\mathbf{g}(\hat{\mathbf{x}}; \xi(\omega)) \leq \mathbf{0} \quad \forall \omega \in G$ , ossia

$$\hat{\mathbf{x}} \in \bigcap_{\omega \in G} \{\mathbf{x} \mid \mathbf{g}(\mathbf{x}; \xi(\omega)) \leq \mathbf{0}\}.$$

**Struttura dell'insieme ammissibile** L'**insieme ammissibile** complessivo può quindi essere scritto come

$$X = \bigcup_{G \in \mathcal{G}} \bigcap_{\omega \in G} \{\mathbf{x} \mid \mathbf{g}(\mathbf{x}; \xi(\omega)) \leq \mathbf{0}\}.$$

Anche se per ogni  $\omega$  l'insieme  $\{\mathbf{x} \mid \mathbf{g}(\mathbf{x}; \xi(\omega)) \leq \mathbf{0}\}$  è convesso, non ci si può aspettare che l'insieme  $X$  sia convesso in generale, poiché l'unione di insiemi convessi non è necessariamente convessa. La convessità del problema complessivo dipende dal tipo di distribuzione (continua o discreta), dalla sua funzione di distribuzione cumulativa, dalla natura dei vincoli (congiunti o disgiunti) e dalla struttura delle funzioni in  $\mathbf{g}$  (lineari, convesse/concave o arbitrarie).

**Vincolo probabilistico lineare con distribuzione normale** Si consideri un vincolo lineare casuale  $\mathbf{a}^T \mathbf{x} \leq b$ , dove  $\mathbf{a} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ , e si richieda  $\mathbb{P}\{\mathbf{a}^T \mathbf{x} \leq b\} \geq \eta$ . Per un dato vettore  $\mathbf{x}$ , la variabile casuale  $\mathbf{a}^T \mathbf{x}$  è distribuita come  $\mathbf{a}^T \mathbf{x} \sim \mathcal{N}(\nu, \sigma^2)$ , dove  $\nu = \boldsymbol{\mu}^T \mathbf{x}$  e  $\sigma^2 = \mathbf{x}^T \Sigma \mathbf{x}$ .

Usando la funzione di distribuzione cumulativa della normale standard  $\Phi(\cdot)$ , il vincolo può essere riscritto come

$$\mathbb{P}\left\{\frac{\mathbf{a}^T \mathbf{x} - \nu}{\sigma} \leq \frac{b - \nu}{\sigma}\right\} = \Phi\left(\frac{b - \nu}{\sigma}\right) \geq \eta \iff \frac{b - \nu}{\sigma} \geq \Phi^{-1}(\eta).$$

Assumendo  $\eta > 0.5$ , quindi  $\Phi^{-1}(\eta) > 0$ , il vincolo diventa

$$\boldsymbol{\mu}^T \mathbf{x} + \Phi^{-1}(\eta) \|\Sigma^{1/2} \mathbf{x}\|_2 \leq b,$$

che è un vincolo conico del secondo ordine. Ne segue che un problema di programmazione lineare con vincoli probabilistici disgiunti di questo tipo è un **problema convesso di tipo SOCP**.

**Convessità della funzione di ricorso** Si consideri ora la funzione di ricorso  $Q(\mathbf{x}) = \mathbb{E}[Q(\mathbf{x}; \boldsymbol{\xi})]$  nel caso di costo di ricorso deterministico, con

$$Q(\mathbf{x}; \boldsymbol{\xi}) = \min_{\mathbf{y}} \left\{ \mathbf{q}^T \mathbf{y} \mid \mathbf{W}\mathbf{y} = \mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x}, \mathbf{y} \geq \mathbf{0} \right\}.$$

Il dominio effettivo di  $Q(\mathbf{x})$  è costituito dai vettori  $\mathbf{x}$  ammissibili per i vincoli di primo stadio tali che  $Q(\mathbf{x}) < +\infty$ , ossia per cui il problema di secondo stadio è (quasi sicuramente) ammissibile.

Per **dualità della programmazione lineare** si ottiene

$$Q(\mathbf{x}; \boldsymbol{\xi}) = \max_{\boldsymbol{\pi}} \left\{ [\mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x}]^T \boldsymbol{\pi} \mid \mathbf{W}^T \boldsymbol{\pi} \leq \mathbf{q} \right\}.$$

Si denoti con  $\Pi = \{\boldsymbol{\pi} \mid \mathbf{W}^T \boldsymbol{\pi} \leq \mathbf{q}\}$  l'**insieme ammissibile del duale**, che è non casuale poiché  $\mathbf{q}$  è deterministico.

**Dimostrazione della convessità** Siano  $\mathbf{x}_1, \mathbf{x}_2 \in D$ ,  $\lambda \in [0, 1]$ , e  $\mathbf{x}_\lambda = \lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$ . Allora Si considerino  $\mathbf{x}^1, \mathbf{x}^2 \in \mathcal{D}$ ,  $\lambda \in [0, 1]$ , e  $\mathbf{x}_\lambda = \lambda\mathbf{x}^1 + (1 - \lambda)\mathbf{x}^2$ .

$$\begin{aligned} Q(\mathbf{x}_\lambda) &\stackrel{\text{def}}{=} \int Q(\mathbf{x}_\lambda, \boldsymbol{\xi}) dP \\ &= \int \max_{\boldsymbol{\pi} \in \Pi} \left\{ [\mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x}_\lambda]^T \boldsymbol{\pi} \right\} dP \\ &= \int \max_{\boldsymbol{\pi} \in \Pi} \left\{ \lambda [\mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x}^1]^T \boldsymbol{\pi} + (1 - \lambda) [\mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x}^2]^T \boldsymbol{\pi} \right\} dP \\ &\leq \lambda \int \max_{\boldsymbol{\pi} \in \Pi} \left\{ [\mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x}^1]^T \boldsymbol{\pi} \right\} dP + (1 - \lambda) \int \max_{\boldsymbol{\pi} \in \Pi} \left\{ [\mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x}^2]^T \boldsymbol{\pi} \right\} dP \\ &= \lambda Q(\mathbf{x}^1) + (1 - \lambda) Q(\mathbf{x}^2). \end{aligned}$$

Pertanto la funzione di ricorso è convessa.

**Osservazioni finali** La dimostrazione precedente risulta particolarmente semplice poiché l'**insieme ammissibile del problema duale è non casuale**. In un contesto più generale, tuttavia, la **convessità della funzione di ricorso** può essere dimostrata anche senza questa assunzione. In particolare, la funzione di ricorso è **tipicamente continua** quando le variabili aleatorie seguono **distribuzioni continue**, mentre nel caso di **distribuzioni discrete** essa assume una struttura **poliedrica**. In entrambi i casi è possibile applicare il **metodo dei piani di taglio di Kelley**. Inoltre, grazie alla **struttura del problema**, tali piani di taglio possono essere generati in modo efficiente tramite **strategie di decomposizione**.

## 2 CAP 2 – Elementi di complessità computazionale

**Recap** La **complessità computazionale** dipende dalla dimensione dell'istanza e può crescere in modo fattoriale, esponenziale o polinomiale. Piccole variazioni nella struttura di un problema possono modificare radicalmente la sua complessità, come nel caso dei problemi di scheduling  $1//L_{\max}$  e  $1/r_i/L_{\max}$ .

La **teoria della NP-completezza** introduce le classi  $P$ ,  $NP$ ,  $NPH$  e  $NPC$  e il concetto di riduzione polinomiale per confrontare la difficoltà dei problemi. I problemi NP-completi sono equivalenti in termini di complessità e l'esistenza di un algoritmo polinomiale per uno di essi implicherebbe  $P = NP$ .

La riduzione da subset-sum mostra che la versione decisionale di  $1/r_i/L_{\max}$  è NP-completa e che il corrispondente problema di ottimizzazione è NP-difficile. **La codifica dei dati è cruciale**: l'algoritmo per il knapsack ha complessità pseudo-polinomiale, poiché dipende dal valore numerico dei dati e non dalla loro lunghezza binaria.

**Complessità di problemi e algoritmi** Si vuole caratterizzare la complessità in funzione della dimensione di un problema. Occorre distinguere la complessità degli algoritmi da quella dei problemi.

Nel caso di un algoritmo caratterizzato da un numero finito di passi, possiamo valutare (eventualmente nel caso peggiore) il **numero di operazioni elementari** in funzione della dimensione  $n$  del problema.

- **Algoritmi enumerativi:** algoritmi che considerano tutte le **permutazioni** di  $n$  oggetti. La loro complessità è  $O(n!)$  ed è certamente non praticabile per valori di  $n$  anche moderatamente grandi.
- **Algoritmi di assegnamento esaustivo:** algoritmi che valutano tutti gli **assegnamenti possibili** di  $n$  variabili binarie. In questo caso la complessità cresce in modo esponenziale ed è pari a  $O(2^n)$ .
- **Algoritmi polinomiali:** un tipico esempio è rappresentato dagli **algoritmi di ordinamento** di  $n$  oggetti. Gli algoritmi più semplici hanno complessità  $O(n^2)$ , mentre algoritmi più efficienti raggiungono complessità  $O(n \log_2 n)$ .

Nel caso di **algoritmi iterativi** che generano una sequenza di soluzioni, si può cercare di caratterizzare la velocità di convergenza (es., lineare o quadratica).

**Complessità intrinseca di un problema** Una questione più sottile si pone quando si vuole caratterizzare la complessità intrinseca di un problema.

Per comprendere la natura della questione, consideriamo un semplice **problema di scheduling di  $n$  job su macchina singola**. Indichiamo con  $p_j$  il tempo necessario per il job  $j = 1, \dots, n$ , e con  $d_j$ ,  $j = 1, \dots, n$  la sua data di consegna (**due date**).

La soluzione è una sequenza di job, ovvero una permutazione  $\sigma$  in cui  $\sigma(k)$  è l'indice del job in posizione  $k$ . I tempi di completamento sono

$$\begin{aligned} C_{\sigma(1)} &= p_{\sigma(1)}, \\ C_{\sigma(k)} &= C_{\sigma(k-1)} + p_{\sigma(k)}, \quad k = 2, \dots, n. \end{aligned}$$

Si vuole minimizzare la massima lateness,

$$L_{\max} \doteq \max_{j \in [n]} L_j,$$

dove  $L_j \doteq C_j - d_j$ . Tale problema viene indicato con la stringa  $1//L_{\max}$ .

**Teorema (regola EDD – Earliest Due Date).** Per il problema  $1//L_{\max}$  esiste una soluzione ottima in cui  $d_{\sigma(k)} \leq d_{\sigma(k+1)}$ .

**Dimostrazione.** Supponiamo che esista una soluzione ottima in cui, per due job consecutivi in sequenza, prima  $i = \sigma(k)$  e poi  $j = \sigma(k+1)$ , si abbia  $d_i > d_j$ . Indichiamo con  $C_i$  e  $C_j = C_i + p_j$  i due tempi di completamento nella soluzione corrente, per la quale abbiamo due valori di lateness  $L_i = C_i - d_i$  e  $L_j = C_j - d_j$ .

Se scambiamo i due job, avremo  $C'_j < C_j$  e  $C'_i = C_j$ . Per il job  $j$ , che anticipiamo, abbiamo  $C'_j < C_j$  e  $L'_j < L_j$ , e quindi la lateness del job  $j$  non può che migliorare nella nuova soluzione. Per il job  $i$ , che viene spostato in avanti nella sequenza, abbiamo

$$L'_i = C'_i - d_i = C_j - d_i < C_j - d_j = L_j,$$

quindi il job  $i$  peggiora la sua lateness, che però non potrà essere peggio della vecchia lateness del job  $j$ . La nuova soluzione migliora quella precedente, contraddicendo l'assunzione di ottimalità.

Abbiamo quindi un algoritmo di complessità polinomiale per il problema. Ma cosa accade se **complichiamo leggermente il problema, introducendo dei tempi di rilascio (release time o ready time)**  $r_i$ ,  $i \in [n]$ , **dei job**? Per il problema  $1/r_i/L_{\max}$  non sono noti algoritmi di complessità polinomiale, ed è facile costruire controsensi alla regola EDD (da adattare comunque alla disponibilità di job).

Esistono algoritmi **branch-and-bound** per il problema  $1/r_i/L_{\max}$  (quindi complessità esponenziale). Non si conosce un algoritmo di complessità polinomiale per questo problema di ottimizzazione combinatoria (e per molti altri), ma neppure è stato dimostrato che esso non possa esistere.

## 2.1 Caratterizzazione della complessità di problemi: classi P, NPC e NPH

Esiste una classe molto ampia di problemi di ottimizzazione per cui non sono disponibili algoritmi di complessità polinomiale. Tuttavia, la questione dell'esistenza o meno di un algoritmo di complessità polinomiale per essi rimane aperta.

La **teoria della NP-completezza** ci permette di dare una risposta parziale alla questione, mostrando come questi problemi siano equivalenti tra di loro, nel senso che un algoritmo di complessità polinomiale per anche uno solo di essi fornirebbe un algoritmo di complessità polinomiale per tutti i problemi di una classe molto ampia.

Il fatto che decenni di ricerca sulla soluzione di tutti questi problemi non abbiano prodotto un algoritmo polinomiale **suggerisce che esso in effetti non esiste**.

Occorre distinguere **problemi di decisione** e **problemi di ottimizzazione**. Esempi di problema di decisione sono i seguenti.

**Problema  $K_0$  (subset sum)** Dati  $n + 1$  numeri interi positivi  $a_1, a_2, \dots, a_n$  e  $b$ , esiste un sottoinsieme  $J \subseteq [n]$  tale che  $\sum_{i \in J} a_i = b$ ?

**Problema  $LS_0$**  Dato un **problema di lot sizing multiprodotto**, con tempi e costi di setup, esiste una soluzione ammissibile rispetto al soddisfacimento della domanda e ai vincoli di capacità produttiva?

Data una specifica istanza di un problema di decisione, la risposta è sì oppure no.

Dal punto di vista teorico, è più agevole trattare problemi di decisione, **ma è facile vedere il legame tra problemi di ottimizzazione e problemi di decisione**. Dato un problema di ottimizzazione  $\min_{x \in S} f(x)$ , **possiamo definirne una versione decisionale**, scegliendo un numero  $k$  e chiedendoci se esiste  $x \in S$  tale che  $f(x) \leq k$ , dove  $k$  è un numero intero. Indichiamo con  $PO$  il problema di ottimizzazione, e con  $PD$  il corrispondente problema di decisione.

Se abbiamo a disposizione un algoritmo efficiente per  $PO$ , allora possiamo risolvere in modo efficiente anche  $PD$ : basta risolvere il problema di ottimizzazione e verificare se  $f(x^*) \leq k$ . Questo ci permette di scrivere  $PD \rightarrow PO$ , nel senso che **il problema di decisione può essere ricondotto al problema di ottimizzazione**.

**Non è detto che tale trasformazione sia conveniente**, ma possiamo escludere che  $PO$  sia facile se  $PD$  è difficile, perché un ipotetico algoritmo efficiente per  $PO$  risolverebbe anche facilmente  $PD$ .

Quindi, per dimostrare che un problema di ottimizzazione è difficile, può bastare dimostrare che è difficile il corrispondente problema di decisione.

D'altro canto, un algoritmo di decisione efficiente potrebbe, in certi casi, essere utilizzato per risolvere il problema di ottimizzazione. Se la funzione di costo in  $PO$  assume valori interi non negativi, e abbiamo un upper bound  $U$  sul costo ottimo, possiamo applicare una procedura di bisezione.

Indichiamo con  $P$  la **classe dei problemi di decisione per cui esiste un algoritmo di complessità polinomiale**, in grado di risolvere tutte le possibili istanze del problema. Con questo vogliamo dire che il numero di passi, e quindi la complessità temporale dell'algoritmo è limitata superiormente da una funzione polinomiale dello spazio di memoria necessario per descrivere ogni istanza del problema. Un tale algoritmo è in grado di fare due cose: **generare una soluzione** e **verificarne la correttezza**. Esistono problemi, come  $K_0$ , per cui la prima parte del compito è difficile, ma la seconda no. Se enumeriamo, mediante un albero di ricerca, tutti i possibili sottoinsiemi  $J$ , possiamo verificare se una specifica istanza ha risposta positiva o negativa, ma chiaramente tale algoritmo ha **complessità esponenziale**.

Tuttavia, se disponessimo di un ipotetico **calcolatore non deterministico**, in grado di eseguire un numero infinito di processi di calcolo in parallelo, saremmo in grado di risolvere il problema in tempo polinomiale.

**Classe NP** Si definisce **classe NP** l'insieme dei problemi di decisione le cui istanze che hanno risposta positiva sono verificabili in tempo polinomiale.

Per definizione,  $P \subseteq NP$ . Una questione meno ovvia è se valga  $P \equiv NP$  o  $P \subset NP$  in senso stretto. È ragionevole, da questo punto di vista, cercare di caratterizzare la sottoclasse dei problemi più difficili in  $NP$ .

**Riduzione in tempo polinomiale** Siano  $P$  e  $Q$  due problemi di decisione, per cui ogni istanza  $I_P$  di  $P$  può essere trasformata in tempo polinomiale in un'istanza  $I_Q$  di  $Q$  tale che  $I_P$  ha risposta positiva se e solo se  $I_Q$  ha risposta positiva. Diremo che  $P$  è riducibile in tempo polinomiale a  $Q$ , e useremo la notazione  $P \prec Q$ .) La notazione  $P \prec Q$  sottolinea che la complessità di  $P$  non è maggiore della complessità di trasformare  $P$  in  $Q$  e poi risolvere  $Q$ ,

$$\text{compl}(P) \leq \text{compl}(Q) + \text{compl}(P \rightarrow Q).$$

Se la trasformazione ha una complessità trascurabile, la riduzione di  $P$  a  $Q$  mostra che  $Q$  non è più facile di  $P$ . Se  $P$  è difficile e  $P \prec Q$ ,  $Q$  non può essere facile. Altrimenti, potremmo trasformare un'istanza di  $P$  in una di  $Q$ , e poi applicare l'algoritmo efficiente per  $Q$ .

**Problemi NP-difficili** Un problema di decisione  $P$  è detto **NP-difficile** (*NP-hard*) se ogni problema nella classe  $NP$  è riducibile a  $P$ . Indichiamo con  $NPH$  la classe dei problemi NP-difficili.

**Problemi NP-completi** Un problema di decisione  $P$  è detto NP-completo se è in  $NP$  ed è NP-hard. Indichiamo con  $NNPC$  la classe dei problemi NP-completi. Le implicazioni pratiche di tali definizioni sono:

1. Un problema NP-difficile non è più facile di un problema qualsiasi in  $NP$ .
2. La classe  $NP$  è la classe dei problemi più difficili in  $NP$ .

Per dimostrare che un problema di decisione  $P$  è NP-completo, occorre dimostrare:

- Che  $P$  è in  $NP$ ;
- Che un problema NP-completo  $Q$  può essere ridotto in tempo polinomiale a  $P$ .

Osserviamo che, dato che  $Q$  è NP-completo,  $P \prec Q$ , e se trascuriamo la complessità della trasformazione, questo implica  $\text{compl}(P) \leq \text{compl}(Q)$ . Ma il secondo passo della dimostrazione di NP-completezza, ovvero dimostrare che  $Q \prec P$ , implica anche che  $\text{compl}(Q) \leq \text{compl}(P)$ . Le due diseguaglianze, sempre a meno della complessità della trasformazione, mostrano che  $\text{compl}(Q) = \text{compl}(P)$ .

In altre parole, **i problemi della classe NPC sono equivalenti in termini di complessità computazionale, e un algoritmo polinomiale per uno di essi permetterebbe di risolverli tutti in tempo polinomiale**. Avremmo quindi  $P = NP$ , ipotesi non troppo plausibile a causa dell'equivalenza di una vasta classe di problemi per i quali non è noto un algoritmo di complessità polinomiale, nonostante essi siano stati oggetto di ampio studio nel corso degli anni.

Se accettiamo l'ipotesi  $P \neq NP$ , possiamo rappresentare le relazioni tra le classi  $P$ ,  $NP$  e  $NP$  come in figura. Essa ipotizza una gerarchia per cui la classe  $P$  sarebbe la classe dei problemi più facili in  $NP$ , e  $NP$  quella dei problemi più difficili.

Tutti i problemi in  $NP$  si possono trasformare nel problema  $Q \in NPC$ . Se  $P \in NP$ , per dimostrare che  $P \in NPC$ , occorre trasformare  $Q$  in  $P$ .

Se dimostriamo che un problema  $P$  è in  $NPC$ , questo può a sua volta essere trasformato in altri problemi, permettendoci di ampliare la classe dei problemi noti in  $NPC$ . Il punto critico, chiaramente, è trovare l'innesco della catena, ovvero il primo problema in  $NPC$ , al quale tutti i problemi in  $NP$  possono essere ricondotti.

Il **teorema di Cook** dimostra che il problema della soddisficiabilità soddisfa i requisiti necessari e ci fornisce la soluzione.

**Teorema di Cook.** Data una formula Booleana in forma canonica disgiuntiva, decidere se esiste un assegnamento di valori ai suoi elementi che la rende vera. Per esempio, la formula  $(A \text{ or } B) \text{ and } (\text{not}(A) \text{ or } C)$ ; definita rispetto alle variabili Booleane  $A$ ,  $B$  e  $C$ , è soddisfatta se  $B$  e  $C$  sono entrambe vere. Al contrario,  $(A \text{ or } B) \text{ and } (\text{not}(A) \text{ or } B) \text{ and } (\text{not}(B))$  non può essere soddisfatta da alcun assegnamento di verità alle variabili.

A partire dal problema della soddisficiabilità, si può ricavare per riduzioni polinomiali successive una famiglia crescente di problemi NP-completi, compreso il problema  $K_0$ , che può essere considerato come un cugino in versione decisionale del problema knapsack.

Il fatto che tale problema faccia parte della classe  $NPC$  ci permette di dimostrare il teorema seguente, che risolve la questione da cui siamo partiti.

**Teorema.** Consideriamo una versione decisionale del problema  $1/r_i/L_{\max}$ : dati i tempi di rilascio  $r_i$ , le date di consegna  $d_i$  e i tempi di lavorazione  $p_i$ , tutti a valori interi positivi, per  $n$  job  $J_i$ ,  $i \in [n]$ , esiste una soluzione in cui nessun job è completato in ritardo? Tale problema di decisione è NP-completo.

**Dimostrazione.** Il problema è chiaramente in  $NP$ , poiché per una data sequenza è facile verificare se i job vengono completati in tempo rispetto alle due date.

Dati gli interi positivi  $a_j$ ,  $j \in [n]$ , creiamo  $n$  job  $J_j$  con parametri

$$r_j = 0, \quad p_j = a_j, \quad d_j = 1 + \sum_{k \in [n]} a_k, \quad j \in [n].$$

Creiamo un ulteriore job  $J_0$  con parametri  $r_0 = b$ ,  $p_0 = 1$ ,  $d_0 = b + 1$ . Perché tutti i job rispettino la data di consegna, è necessario che il job  $J_0$  inizi la lavorazione al tempo  $t = b$ . Inoltre, dato che la data di consegna degli altri job è pari alla somma di tutti i tempi di lavorazione, la soluzione non può presentare periodi di tempo in cui la macchina è ferma, prima di avere completato l'intero insieme di job.

Questo richiede che sia possibile individuare un sottoinsieme  $J$  di job da schedulare prima di  $J_0$ , in modo tale che

$$\sum_{j \in J} p_j = r_0.$$

Tale insieme risolve il problema *subset-sum*.

## 2.2 Dai problemi di decisione ai problemi di ottimizzazione

Il teorema dimostra che un problema di decisione legato al problema di ottimizzazione  $1/r_i/L_{\max}$  è NP-completo, **ma cosa possiamo dire del problema di ottimizzazione stesso?**

Per definizione, la classe  $NPC$  contiene solo problemi di decisione. Possiamo però estendere le classi  $P$  e  $NPH$ , includendo in esse anche problemi di ottimizzazione.

I **problematici di ottimizzazione** per cui è noto un algoritmo di complessità polinomiale stanno in  $P$ . Nella classe  $NPH$  possiamo includere problemi di ottimizzazione ai quali possiamo ricondurre un corrispondente problema di decisione. **La versione decisionale di  $1/r_i/L_{\max}$  si riduce chiaramente al problema di ottimizzazione.**

Inoltre, nella dimostrazione abbiamo assunto che i dati fossero numeri interi. Ma il problema a numeri interi può evidentemente essere ridotto al problema generale. Possiamo quindi affermare che il problema di scheduling  $1/r_i/L_{\max}$  è NP-difficile.

**L'impatto della codifica di un problema** Nella trattazione ci siamo limitati alle classi fondamentali e siamo stati piuttosto informali e imprecisi. Non possiamo però fare a meno di considerare l'impatto del modo in cui si codifica un problema. Sarebbe infatti errato, per esempio, associare a un problema knapsack una dimensione pari al numero di oggetti. La dimensione si riferisce a una codifica binaria che comprende tutti i dati del problema. Per mostrare la rilevanza di ciò, **consideriamo un classico algoritmo di programmazione dinamica** per la soluzione del problema knapsack:

$\begin{aligned} \max \quad & \sum_{k=1}^n v_k x_k \\ \text{s.t.} \quad & \sum_{k=1}^n w_k x_k \leq B \\ & x_k \in \{0, 1\}, \quad k = 1, \dots, n. \end{aligned}$
--

Definiamo la funzione valore

$$V_k(s) := \text{valore del sottoinsieme ottimale tra gli oggetti } \{k, k+1, \dots, n\},$$

quando la capacità residua è  $s$ . In sostanza, la funzione valore assume che siano già state fatte scelte di inserimento o meno degli oggetti da 1 a  $k-1$ ; a valle di tale selezione, abbiamo una capacità residua  $s$ , e ci chiediamo come utilizzarla al meglio per le scelte rimanenti. Se i dati  $w_k$  e  $B$  del problema sono interi, lo sarà anche la capacità residua  $s$ .

Per risolvere il problema, ovvero trovare il valore di  $V_1(B)$ , si applica una relazione ricorsiva:

$$V_k(s) = \begin{cases} V_{k+1}(s), & 0 \leq s < w_k, \\ \max\{V_{k+1}(s), V_{k+1}(s - w_k) + v_k\}, & w_k \leq s \leq B. \end{cases}$$

Si tratta di una equazione funzionale con condizione terminale:

$$V_n(s) = \begin{cases} 0, & 0 \leq s < w_n, \\ v_n, & w_n \leq s \leq B. \end{cases}$$

Occorre tabulare tutte le funzioni  $V_k(s)$ ,  $k = 1, \dots, n$ , per valori interi di  $s$ , che assume valori nel range da 0 a  $B$ . Pertanto, tale algoritmo ha complessità  $O(nB)$ .

Questo non implica che il problema knapsack abbia complessità polinomiale: per rappresentare il valore  $B$  in aritmetica binaria bastano  $\log_2 B$  bit. Quindi **l'algoritmo, rispetto a tale codifica binaria, ha complessità esponenziale**. Se si utilizzasse un **computer con una codifica unaria**, l'algoritmo che abbiamo considerato avrebbe **complessità polinomiale**. Si dice infatti che un algoritmo di questo tipo è *pseudo-polynomial*.

## 3 CAP 3 –Metodi di decomposizione in ottimizzazione

### 3.1 Motivation

**Ruolo dei metodi di decomposizione** I metodi di **decomposizione** svolgono un ruolo centrale nell'ottimizzazione, poiché consentono:

- di sfruttare **strutture favorevoli** del problema, come:
  - sottoproblemi di rete;
  - trasformazioni da problemi **NP-hard** a problemi risolvibili in **tempo polinomiale**;
- di **parallelizzare** la soluzione di problemi di grande scala;
- di affrontare modelli di **ottimizzazione stocastica** di grandi dimensioni basati su scenari;
- di trattare problemi combinatori difficili mediante una **sequenza di sottoproblemi più semplici**, eventualmente combinando diverse strategie di soluzione;
- di evitare difficoltà di modellazione dovute a **vincoli molto complessi**.

### 3.2 Problemi separabili

**Problema separabile** Si consideri il problema di ottimizzazione

$$\text{opt} \sum_{j \in [n]} f_j(x_j) \quad \text{s.t. } x_j \in S_j.$$

In questo caso il problema può essere **decomposto** in sottoproblemi indipendenti

$$\text{opt}_{x_j \in S_j} f_j(x_j),$$

poiché i sottovettori  $x_j$  sono soggetti a **vincoli indipendenti**.

**Interpretazione nei modelli di programmazione lineare** Nel caso di un modello di programmazione lineare

$$\min c^T x \quad \text{s.t. } Ax = b, \quad x \geq 0,$$

la separabilità corrisponde a una struttura **a blocchi diagonali** della matrice tecnologica:

$$A = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \\ \vdots \\ D_n \end{bmatrix}.$$

### 3.3 Fattori complicanti

**Assenza di struttura completamente separabile** Nella maggior parte dei casi reali, non si dispone di una struttura così favorevole e sono presenti **fattori complicanti**.

**Struttura block-angular** Un caso tipico è quello di un problema LP di grande scala con struttura **block-angular**:

$$A = \begin{bmatrix} C_1 & C_2 & C_3 & \cdots & C_n \\ D_1 & & & & \\ D_2 & & & & \\ D_3 & & & & \\ \vdots & & & & \\ D_n & & & & \end{bmatrix}.$$

In questo caso il fattore complicante è rappresentato da un insieme di **vincoli di interazione** che accoppiano i sottoproblemi e impediscono la decomposizione basata sulla struttura a blocchi diagonali.

**Rilassamento dei vincoli di interazione** La decomposizione diventa possibile se i vincoli di interazione vengono **rilassati** in qualche modo, ad esempio tramite la **decomposizione lagrangiana duale**.

**Variabili di interazione** Un secondo caso è quello in cui la matrice ha la forma

$$A = \begin{bmatrix} C_1 & D_1 \\ C_2 & D_2 \\ C_3 & D_3 \\ \vdots & \vdots \\ C_n & D_n \end{bmatrix}.$$

Qui il fattore complicante è costituito da **variabili di interazione**, che impediscono la decomposizione.

**Fissaggio delle variabili** La decomposizione è possibile se si **fissano** le variabili di interazione, come avviene nella **decomposizione L-shaped**, che corrisponde alla **decomposizione di Benders** nella programmazione stocastica.

### 3.4 Panoramica dei metodi di decomposizione

**Classificazione dei metodi** Esiste un ampio insieme di metodi di decomposizione tra loro correlati. Alcuni sono **esatti**, almeno in linea di principio, mentre altri conducono ad **algoritmi approssimati**:

- rilassamento lagrangiano e decomposizione lagrangiana;
- euristiche duali;

- decomposizione di Dantzig–Wolfe;
- generazione di colonne;
- decomposizione gerarchica;
- matheuristics (da non confondere con le metaeuristiche);
- decomposizione di Benders per MILP con struttura speciale;
- decomposizione L-shaped nella programmazione stocastica a due stadi con ricorso;
- progressive hedging per la programmazione stocastica multistadio con ricorso;
- programmazione dinamica (decomposizione basata sul tempo).

## 3.5 Part 1 – Dual decomposition and progressive hedging for stochastic programming

### 3.5.1 Dual decomposition

**Problema di partenza** Per comprendere come la dualità possa essere utilizzata in un contesto semplice, si consideri il problema:

$$\max \sum_{i=1}^n f_i(x_i) \quad (1)$$

$$\text{s.t. } \sum_{i=1}^n g_i(x_i) \leq b, \quad (2)$$

$$x_i \in S_i, \quad i = 1, \dots, n. \quad (3)$$

Le variabili decisionali  $x_i, i = 1, \dots, n$ , possono essere interpretate come **attività** che generano un profitto  $f_i(x_i)$  e consumano una quantità di risorsa  $g_i(x_i)$ . La funzione obiettivo (1) rappresenta il **profitto totale**, mentre il vincolo (2) è un **vincolo di budget** sulla risorsa.

**Osservazione sulle unità di misura** La funzione obiettivo è misurata in termini monetari, mentre il parametro  $b$  è espresso in unità di risorsa. Se fosse possibile eliminare il vincolo di budget, il problema risulterebbe immediatamente **decomponibile**.

**Dualizzazione del vincolo di budget** Il vincolo di budget viene dualizzato introducendo un moltiplicatore di Lagrange  $\mu \geq 0$  e scrivendo la funzione lagrangiana:

$$L(x; \mu) = \sum_{i=1}^n f_i(x_i) + \mu \left( b - \sum_{i=1}^n g_i(x_i) \right) = \sum_{i=1}^n [f_i(x_i) - \mu g_i(x_i)] + \mu b.$$

**Problemi primali decompositi** La funzione lagrangiana deve essere massimizzata rispetto alle variabili primarie, ottenendo un insieme di sottoproblemi indipendenti:

$$\max_{x_i \in S_i} [f_i(x_i) - \mu g_i(x_i)] \equiv \pi_i(x_i).$$

Ogni sottoproblema consiste nel massimizzare il contributo di profitto netto, dato dal profitto meno il costo della risorsa.

**Interpretazione economica del moltiplicatore** Il moltiplicatore  $\mu$  rappresenta un **prezzo ombra** della risorsa, misurato in unità monetarie per unità di risorsa. In questo caso, il problema duale consiste nel **minimizzare** la funzione duale rispetto a  $\mu$ .

**Sottogradiente della funzione duale** Date le soluzioni rilassate  $x_i^*$ , un sottogradiente della funzione duale è:

$$\sum_{i=1}^n g_i(x_i^*) - b.$$

Questo valore è positivo quando il budget viene superato; in tal caso il prezzo della risorsa deve essere aumentato. Il prezzo deve invece essere ridotto quando il budget non viene completamente utilizzato.

**Schema domanda–offerta** Il meccanismo risultante può essere interpretato come uno schema di **domanda–offerta**, in cui il prezzo della risorsa viene aggiornato in funzione dello squilibrio tra consumo e disponibilità.

**Osservazioni sulla dual decomposition** La decomposizione duale può convergere lentamente nella pratica, ma può risultare efficace per alcuni problemi di grande scala con struttura particolare. Talvolta è sufficiente ottenere una soluzione di buona qualità: se da una decomposizione duale è possibile ricostruire una soluzione primale ammissibile, si ottiene una **dual heuristic**.

**Lagrangiano aumentato** I metodi lagrangiani possono essere integrati con metodi a funzione di penalità, ottenendo schemi di **Lagrangiano aumentato**, ad esempio minimizzando:

$$f(x) + \sum_{i \in I} \lambda_i h_i(x) + \sigma \sum_{i \in I} h_i^2(x),$$

nel caso di problemi con vincoli di uguaglianza.

### 3.5.2 Progressive hedging for multistage stochastic programming

**Rappresentazione dell’incertezza** Nella programmazione stocastica multistadio, l’incertezza è rappresentata tramite **alberi di scenario**. In una formulazione compatta del modello, si definiscono variabili decisionali  $x_n$  associate ai nodi  $n$  dell’albero.

**Formulazione split-variable** In una formulazione a variabili separate, si introducono variabili  $x_t^s$ , associate allo scenario  $s$  al tempo  $t$ .

**Vincoli di non anticipatività** In questo caso è necessario imporre esplicitamente i **vincoli di non anticipatività**: variabili decisionali corrispondenti a scenari diversi allo stesso tempo  $t$  devono assumere lo stesso valore se gli scenari sono indistinguibili a quel tempo.

**Esempio di non anticipatività** Nel caso illustrato, i nodi attraversati dagli scenari  $s = 1, 2, 3, 4$  al tempo  $t = 1$  sono indistinguibili; pertanto:

$$x_{1i}^1 = x_{1i}^2 = x_{1i}^3 = x_{1i}^4, \quad i = 1, \dots, I.$$

Analogamente, al tempo  $t = 2$  si hanno vincoli come:

$$x_{2i}^5 = x_{2i}^6, \quad i = 1, \dots, I.$$

**Insiemi di scenari indistinguibili** Si denoti con  $\mathcal{F}_s^t$  l’insieme degli scenari indistinguibili da  $s$  fino al tempo  $t$ . Ad esempio:

$$\mathcal{F}_1^0 = \{1, 2, 3, 4, 5, 6, 7, 8\}, \quad \mathcal{F}_2^1 = \{1, 2, 3, 4\}, \quad \mathcal{F}_5^2 = \{5, 6\}.$$

I vincoli di non anticipatività possono essere scritti come:

$$x_{it}^s = x_{it}^{s'}, \quad \forall i, t, s, s' \in \mathcal{F}_s^t.$$

**Formulazione del problema multistadio** Si consideri un sistema dinamico con stato  $z_t$  e equazione di stato:

$$z_{t+1} = G_t(z_t, x_t, \xi_{t+1}), \quad t = 0, 1, \dots, T,$$

dove  $x_t$  è la variabile di controllo e  $\xi_{t+1}$  è una variabile aleatoria.

Sia  $s = (\xi_1^s, \xi_2^s, \dots, \xi_{T+1}^s)$  uno scenario con probabilità  $\pi_s$ . Il problema di scenario individuale è:

$$\min \sum_{t=0}^T \gamma^t f_t(z_t, x_t, \xi_{t+1}^s) + \gamma^{T+1} Q(z_{T+1})$$

soggetto a:

$$\begin{aligned} z_{t+1} &= G_t(z_t, x_t, \xi_{t+1}^s), \quad t = 0, 1, \dots, T, \\ L_t(z_t) &\leq x_t \leq U_t(z_t), \quad t = 0, 1, \dots, T. \end{aligned}$$

**Problema stocastico con non anticipatività** Il problema complessivo può essere scritto come:

$$\min \sum_{s \in S} \pi_s \left( \sum_{t=0}^T \gamma^t f_t(z_t^s, x_t^s, \xi_{t+1}^s) + \gamma^{T+1} Q(z_{T+1}^s) \right)$$

soggetto a:

$$\begin{aligned} z_{t+1}^s &= G_t(z_t^s, x_t^s, \xi_{t+1}^s), \\ L_t(z_t^s) &\leq x_t^s \leq U_t(z_t^s), \\ x_t^s &= \frac{\sum_{s' \in \mathcal{F}_s^t} \pi_{s'} x_t^{s'}}{P(\mathcal{F}_s^t)}, \quad \forall s, t. \end{aligned} \tag{4}$$

**Dualizzazione e regolarizzazione** Introducendo moltiplicatori di Lagrange  $w_t^s$  per i vincoli di non anticipatività e dualizzando, il problema può essere decomposto per scenario. Per migliorare la convergenza si introduce un termine di penalità quadratico, ottenendo per ogni scenario:

$$\min \sum_{t=0}^T \gamma^t \left\{ f_t(z_t^s, x_t^s, \xi_{t+1}^s) + w_t^s x_t^s + \frac{\rho}{2} (x_t^s - x(\mathcal{F}_s^t))^2 \right\} + \gamma^{T+1} Q(z_{T+1}^s).$$

**Osservazioni finali** Il termine quadratico può essere interpretato come una **regolarizzazione**. La scelta del parametro  $\rho$  e della strategia di aggiornamento dei moltiplicatori è critica per la convergenza. Nonostante ciò, il **progressive hedging** è uno strumento utile per sfruttare la struttura del problema ed è stato utilizzato per costruire euristiche per problemi stocastici multistadio misti interi difficili.

## 3.6 Part 2 – Benders decomposition and L-shaped decomposition for stochastic programming

### 3.6.1 Background: Kelley's cutting planes

**Problema convesso e iperpiani di supporto** Si consideri il problema convesso

$$\min_{x \in S} f(x),$$

e si supponga che, dato un punto  $x^k$ , sia possibile calcolare non solo il valore della funzione  $f(x^k) = \alpha_k$ , ma anche un **subgradiente**  $\gamma_k$ , che esiste se la funzione è convessa sull'insieme  $S$ .

In altre parole, è possibile trovare una funzione affine tale che

$$f(x^k) = \alpha_k + \gamma_k^T x^k \tag{5}$$

$$f(x) \geq \alpha_k + \gamma_k^T x \quad \forall x \in S. \tag{6}$$

**Approssimazione dal basso** La disponibilità di un tale iperpiano di supporto suggerisce la possibilità di approssimare  $f$  dal basso, tramite l'**inviluppo superiore** degli iperpiani di supporto.

**Idea dell'algoritmo di Kelley** L'algoritmo di **Kelley cutting planes** sfrutta questa idea costruendo e migliorando una funzione di bound inferiore fino a soddisfare un criterio di convergenza. Se  $S$  è **poliedrale**, occorre risolvere una sequenza di **LP**.

### 3.6.2 Kelley's cutting planes come algoritmo

**Algoritmo (cutting planes di Kelley)** **Step 0.** Sia  $x^1 \in S$  una soluzione ammissibile iniziale; inizializzare il contatore delle iterazioni  $k \leftarrow 0$ , l'upper bound  $u_0 = f(x^1)$ , il lower bound  $l_0 = -1$ , e la funzione di bound inferiore  $\phi_0(x) = -1$ .

**Step 1.** Incrementare il contatore  $k \leftarrow k + 1$ . Trovare un subgradiente di  $f$  in  $x^k$  tale che (5) e (6) siano soddisfatte.

**Step 2.** Aggiornare l'upper bound

$$u_k = \min\{u_{k-1}, f(x^k)\},$$

e la funzione di bound inferiore

$$\phi_k(x) = \max\{\phi_{k-1}(x), \alpha_k + \gamma_k^T x\}.$$

**Step 3.** Risolvere il problema

$$l_k = \min_{x \in S} \phi_k(x),$$

e sia  $x^{k+1}$  la soluzione ottima.

**Step 4.** Se  $u_k - l_k < \varepsilon$ , arrestare:  $x^{k+1}$  è un'approssimazione soddisfacente della soluzione ottima; altrimenti, tornare allo Step 1.

### 3.6.3 Convexity properties of stochastic programs

**Due-stadi SLP con ricorso** Si consideri il problema SLP a due stadi con ricorso:

$$\min c^T x + \mathbb{E}_\xi [Q(x; \xi)]$$

$$\text{s.t. } Ax = b, \quad x \geq 0,$$

dove

$$Q(x; \xi) \equiv \min_y \{q(\xi)^T y \mid W(\xi)y = h(\xi) - T(\xi)x, y \geq 0\}.$$

Si parla di **fixed recourse** quando  $W$  è deterministica, mentre si parla di **random recourse** nel caso generale.

**Funzione di ricorso con costo deterministico** Si consideri la funzione di ricorso  $Q(x) \equiv \mathbb{E}[Q(x; \xi)]$  nel caso di costo di ricorso deterministico:

$$Q(x; \xi) \equiv \min_y \{q^T y \mid Wy = h(\xi) - T(\xi)x, y \geq 0\}.$$

**Dominio effettivo** Il dominio effettivo di  $Q(x)$  consiste dei vettori  $x$  che sono ammissibili per i vincoli di primo stadio e tali che  $Q(x) < +\infty$ , cioè per cui il problema di secondo stadio è (quasi sicuramente) ammissibile.

**Nota sulle ipotesi** Si assume che, esclusa l'inammissibilità del secondo stadio, l'aspettazione sia sempre definita, cioè che le variabili aleatorie non siano heavy-tailed.

**Dualità LP e forma duale della funzione di ricorso** Per dualità LP,

$$Q(x; \xi) = \max_{\pi} \left\{ [h(\xi) - T(\xi)x]^T \pi \mid W^T \pi \leq q \right\}.$$

Si denoti l'insieme ammissibile del duale con

$$\Gamma = \{\pi \mid W^T \pi \leq q\}.$$

Poiché  $q$  è deterministico, tale insieme è **non aleatorio**.

**Dimostrazione di convessità (caso con regione duale non aleatoria)** Siano  $x_1, x_2 \in D$ ,  $\lambda \in [0, 1]$  e  $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$ . Allora:

$$\begin{aligned} Q(x_\lambda) &= \int_{\Omega} Q(x_\lambda; \xi) dP = \int_{\Omega} \max_{\pi \in \Gamma} \left\{ [h(\xi) - T(\xi)x_\lambda]^T \pi \right\} dP \\ &= \int_{\Omega} \max_{\pi \in \Gamma} \left\{ \lambda [h(\xi) - T(\xi)x_1]^T \pi + (1 - \lambda) [h(\xi) - T(\xi)x_2]^T \pi \right\} dP \\ &\leq \lambda \int_{\Omega} \max_{\pi \in \Gamma} \left\{ [h(\xi) - T(\xi)x_1]^T \pi \right\} dP + (1 - \lambda) \int_{\Omega} \max_{\pi \in \Gamma} \left\{ [h(\xi) - T(\xi)x_2]^T \pi \right\} dP \\ &= \lambda Q(x_1) + (1 - \lambda) Q(x_2). \end{aligned}$$

**Convexity of the recourse function** La dimostrazione precedente è semplice perché la regione ammissibile duale è non aleatoria. In generale, la convessità della funzione di ricorso può essere mostrata anche in contesti più generali. In sintesi:

- la funzione di ricorso è tipicamente **differenziabile** per distribuzioni continue;
- la funzione di ricorso è **poliedrale** per distribuzioni discrete.

In entrambi i casi, si può fare affidamento sui **cutting planes di Kelley** per risolvere il problema. Dato lo **schema di decomposizione** del problema, i cutting planes possono essere ottenuti tramite una strategia di decomposizione.

### 3.6.4 L-shaped decomposition

**Deterministic equivalent** Si consideri l'equivalente deterministico:

$$\begin{aligned} \min \quad & c^T x + \sum_{s \in S} p_s q^T y^s \\ \text{s.t.} \quad & Ax = b, \\ & Wy^s + T_s x = h_s \quad \forall s \in S, \\ & x, y^s \geq 0. \end{aligned}$$

La matrice tecnologica del problema complessivo ha una struttura **dual block-angular**:

$$\begin{bmatrix} A & 0 & 0 & \cdots & 0 \\ T_1 & W & 0 & \cdots & 0 \\ T_2 & 0 & W & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T_S & 0 & 0 & \cdots & W \end{bmatrix}.$$

Dato un primo stadio fissato, si ottengono problemi di secondo stadio **indipendenti per scenario**.

**Riformulazione con variabile ausiliaria** Si riscrive il problema a due stadi come:

$$\begin{aligned} & \min c^T x + \theta \\ & \text{s.t. } Ax = b, \\ & \quad \theta \geq Q(x) \\ & \quad x \geq 0. \end{aligned} \tag{7}$$

Si costruisce un **master problem rilassato** rilassando (7), che viene approssimato tramite cutting planes:

- **optimality cuts** della forma  $\theta \geq \alpha^T x + \beta$ ;
- **feasibility cuts** della forma  $0 \geq \alpha^T x + \beta$ .

I coefficienti di ciascun taglio si ottengono risolvendo i sottoproblemi di scenario per una data decisione di primo stadio.

**L-shaped decomposition: optimality cuts** Sia  $\hat{x}$  la soluzione ottima del master iniziale e si consideri il duale del problema di secondo stadio per lo scenario  $s$ :

$$Q_s(\hat{x}) \equiv \max(h_s - T_s \hat{x})^T \pi^s \quad \text{s.t. } W^T \pi^s \leq q_s. \tag{8}$$

Data una soluzione duale ottima  $\hat{\pi}^s$ , valgono:

$$Q_s(\hat{x}) = (h_s - T_s \hat{x})^T \hat{\pi}^s \tag{9}$$

$$Q_s(x) \geq (h_s - T_s x)^T \hat{\pi}^s \quad \forall x. \tag{10}$$

La (10) deriva dal fatto che  $\hat{\pi}^s$  è ottima per  $\hat{x}$ , ma non necessariamente per un generico  $x$ .

Sommando (10) sugli scenari:

$$Q(x) = \sum_{s \in S} p_s Q_s(x) \geq \sum_{s \in S} p_s (h_s - T_s x)^T \hat{\pi}^s.$$

Quindi si aggiunge al master rilassato il seguente **optimality cut**:

$$\theta \geq \sum_{s \in S} p_s (h_s - T_s x)^T \hat{\pi}^s.$$

**L-shaped decomposition: feasibility cuts** Se il ricorso non è completo, alcuni sottoproblemi di scenario possono essere inammissibili per una decisione di primo stadio  $\hat{x}$ . Si può allora sfruttare il duale del sottoproblema di scenario per trovare un **feasibility cut** che elimina  $\hat{x}$  da ulteriori considerazioni.

Si osservi che la regione ammissibile del duale non dipende dalle decisioni di primo stadio, poiché  $\hat{x}$  non compare nei vincoli (8). Se il duale è inammissibile, allora il problema di secondo stadio dello scenario corrispondente sarebbe inammissibile per qualsiasi decisione di primo stadio; questo caso viene escluso e può indicare un errore di modellazione.

Quando il primale è inammissibile, il duale è illimitato. Esiste dunque un raggio estremo della regione duale lungo il quale la soluzione duale tende a infinito. Il duale del secondo stadio è illimitato se esistono variabili duali  $\pi^*$  tali che

$$W^T \pi^* \leq 0, \quad (h_s - T_s \hat{x})^T \pi^* > 0.$$

La fattibilità del primale richiede

$$W y = h_s - T_s x \Rightarrow (\pi^*)^T W y = (\pi^*)^T (h_s - T_s x),$$

e, usando  $W^T \pi^* \leq 0$  e  $y \geq 0$ , si ottiene

$$(\pi^*)^T (h_s - T_s x) \leq 0.$$

Quindi

$$(\pi^*)^T (h_s - T_s x) \leq 0$$

è un taglio valido che viene aggiunto al master rilassato.

**Schema iterativo dell'L-shaped decomposition** L-shaped decomposition alterna:

1. la soluzione del master problem rilassato, che produce  $\hat{\theta}$  e  $\hat{x}$ ;
2. la soluzione dei corrispondenti sottoproblemi di scenario.

A ogni iterazione, vengono aggiunti tagli al master. L'algoritmo si ferma quando la soluzione ottima del master soddisfa

$$\hat{\theta} \geq Q(\hat{x}).$$

Questa condizione può essere rilassata se è sufficiente una soluzione near-optimal.

**L-shaped decomposition: variations on the theme** Esistono varianti per gestire i tagli e migliorare la convergenza (ad esempio **regularized L-shaped decomposition**). L'idea può essere generalizzata a problemi multistadio:

- nested Benders decomposition;
- abridged nested Benders decomposition;
- stochastic dual dynamic programming.

Nei metodi di decomposizione stocastica (Higle e Sen) si ha un'interazione tra campionamento e ottimizzazione: l'idea è costruire tagli asintoticamente validi. Strategie alternative si basano sul vincolare la funzione di ricorso superiormente e inferiormente, partizionando il dominio per ottenere bound stretti. Quando metodi del simplesso o interior point riescono a risolvere un LP stocastico, spesso non c'è vantaggio nell'uso della decomposizione; tuttavia, la decomposizione è necessaria per problemi realmente di grande scala.

## 3.7 Part 3 – Dantzig–Wolfe decomposition and column generation

### 3.7.1 Cutting stock by column generation

**Problema di cutting stock** Si devono tagliare rotoli (tutti di larghezza  $L$ ) di un certo materiale in rotoli più corti, in modo da soddisfare una domanda assegnata. Si introduca un insieme di  $n$  item con domanda  $d_i$  e larghezza  $w_i \leq L$ ,  $i \in [n]$ . L'obiettivo è impacchettare gli item in modo da utilizzare il **numero minimo di rotoli**. Questo è un esempio stilizzato di una vasta classe di problemi noti come **cutting-stock problems**. :contentReference[oaicite:1]index=1

**Modello MILP con variabili per rotolo** Siano disponibili  $m$  rotoli, indicizzati da  $j \in [m]$ , e si introducano le variabili decisionali:

- $\delta_j \in \{0, 1\}$ , uguale a 1 se si usa il rotolo  $j$ ;
- $x_{ij} \in \mathbb{Z}_+$ , numero di item  $i$  tagliati dal rotolo  $j$ .

Un possibile modello MILP è:

$$\begin{aligned} & \min \sum_{j \in [m]} \delta_j \\ \text{s.t. } & \sum_{i \in [n]} w_i x_{ij} \leq L \delta_j, \quad j \in [m] \\ & \sum_{j \in [m]} x_{ij} = d_i, \quad i \in [n] \\ & \delta_j \in \{0, 1\}, \quad x_{ij} \in \mathbb{Z}_+, \quad i \in [n], \quad j \in [m]. \end{aligned}$$

Pur essendo corretto dal punto di vista teorico, il modello soffre di difficoltà legate a vincoli **big-M** e a problemi di **simmetria** (molte soluzioni ottime equivalenti possono penalizzare i solver commerciali). :contentReference[oaicite:2]index=2

**Modello a pattern di taglio** Si passa a una formulazione basata su un insieme di **cutting patterns**. Ogni pattern, indicizzato da  $k$ , è associato al numero  $a_{ik}$  di item  $i$  ottenuti da un rotolo. Dato un insieme  $K$  di pattern, si introduce il numero intero  $\gamma_k$  di rotoli tagliati secondo il pattern  $k$ :

$$\begin{aligned} & \min \sum_{k \in K} \gamma_k \\ \text{s.t. } & \sum_{k \in K} a_{ik} \gamma_k = d_i, \quad i \in [n] \\ & \gamma_k \in \mathbb{Z}_+, \quad k \in K. \end{aligned}$$

Questa formulazione risolve la simmetria, ma la cardinalità di  $K$  è enorme: solo pochi pattern sono davvero utili e non è ovvio come restringere  $K$  *a priori*. Si parte quindi con un numero limitato di pattern e li si genera dinamicamente tramite **column generation**.

### 3.7.2 Generating a new column by pricing

**LP-rilassato del modello ristretto** Si consideri un LP-rilassato di una versione ristretta del modello:

$$\begin{aligned} & \min \sum_{k \in K'} \gamma_k \\ \text{s.t. } & \sum_{k \in K'} a_{ik} \gamma_k = d_i, \quad i \in [n] \\ & \gamma_k \geq 0, \quad k \in K', \end{aligned} \tag{11}$$

dove  $K' \subset K$ . Come migliorare la soluzione introducendo una nuova colonna? :contentReference[oaicite:4]index=

**Costo ridotto e scelta della colonna entrante** Siano  $\pi_i$ ,  $i \in [n]$ , le variabili duali dei vincoli (11). Dalla teoria del simplex primale, conviene introdurre in base una variabile con **costo ridotto negativo**. Per un pattern alternativo  $k$ , il costo ridotto è:

$$\hat{c}_k = 1 - \sum_{i \in [n]} \pi_i a_{ik}.$$

Si cerca la colonna  $q$  con costo ridotto minimo,  $\hat{c}_q = \min_k \hat{c}_k$ . Se  $\hat{c}_q \geq 0$ , la base corrente è ottima; altrimenti si introduce la variabile non basica  $\gamma_q$  e si itera. :contentReference[oaicite:5]index=5

**Pricing come knapsack** La minimizzazione si ottiene risolvendo un problema intero di knapsack:

$$\begin{aligned} & \max \sum_{i \in [n]} \pi_i y_i \\ \text{s.t. } & \sum_{i \in [n]} w_i y_i \leq L \\ & y_i \in \mathbb{Z}_+. \end{aligned}$$

Questo produce un nuovo pattern, che viene introdotto se il valore ottimo è maggiore di 1.

**Osservazioni su branch-and-price e complessità** Questo fornisce un lower bound per un algoritmo **branch-and-price**. In alternativa, si può generare un insieme di colonne e poi ripristinare l'integralità delle variabili  $\gamma_k$  risolvendo un singolo IP (euristica). In casi meno stilizzati, la complessità dei vincoli può essere “nascosta” nel sottoproblema di pricing. Lo schema di column generation è al cuore della **decomposizione di Dantzig–Wolfe**.

### 3.7.3 Dantzig–Wolfe decomposition

**Modello LP e separazione dei vincoli** Si consideri un modello LP:

$$\min c^T x$$

$$\text{s.t. } Ax \geq b \tag{12}$$

$$Dx \geq f \tag{13}$$

$$x \geq 0_n.$$

Si separano i vincoli interpretando (12) come vincoli **complicanti**, mentre (13) (insieme alla non-negatività) come vincoli **facili**.

**Poliedro facile e Minkowski–Weyl** Si definisce il poliedro:

$$X := \{x \in \mathbb{R}^n \mid Dx \geq f, x \geq 0_n\}.$$

Si assume che minimizzare una funzione lineare su  $X$  sia facile (ad esempio perché  $D$  è block-diagonal o per struttura favorevole come un network flow). Per il teorema di Minkowski–Weyl,  $X$  si può rappresentare come somma di un politopo e di un cono:

$$X = \text{conv}\{V\} + \text{cone}\{D\},$$

dove  $V = \{v^1, \dots, v^Q\}$  è l'insieme dei punti estremi e  $D = \{d^1, \dots, d^R\}$  l'insieme delle direzioni estreme.

**Rappresentazione mediante combinazioni convesse e coniche** Il vincolo  $x \in X$  si esprime come:

$$x = \sum_{q \in V} \lambda_q v^q + \sum_{r \in D} \mu_r d^r \tag{14}$$

$$\sum_{q \in V} \lambda_q = 1 \tag{15}$$

$$\lambda_q \geq 0, q \in V; \quad \mu_r \geq 0, r \in D. \tag{16}$$

dove (15) è il **vincolo di convessità**. Sostituendo (14) nel modello originale si ottiene un problema in  $\lambda_q$  e  $\mu_r$ .

**Master Problem (MP)** Definendo gli scalari e vettori:

$$c_q := c^T v^q, \quad a_q := Av^q, \quad q \in V,$$

$$c_r := c^T d^r, \quad a_r := Ad^r, \quad r \in D,$$

si ottiene il **master problem** (MP):

$$(MP) \quad \min \sum_{q \in V} c_q \lambda_q + \sum_{r \in D} c_r \mu_r$$

$$\text{s.t. } \sum_{q \in V} a_q \lambda_q + \sum_{r \in D} a_r \mu_r \geq b$$

$$\sum_{q \in V} \lambda_q = 1$$

$$\lambda_q \geq 0, q \in V; \quad \mu_r \geq 0, r \in D.$$

Il problema è che MP può coinvolgere un numero enorme di variabili e non è in generale possibile enumerare tutti i punti/direzioni estremi di  $X$  a partire dalla descrizione per diseguaglianze.

**Restricted Master Problem (RMP)** La strategia si basa su **column generation**. Si inizializza un Restricted Master Problem con un sottoinsieme  $V' \subset V$  di punti estremi e  $D' \subset D$  di direzioni estreme:

$$(RMP) \quad \min \sum_{q \in V'} c_q \lambda_q + \sum_{r \in D'} c_r \mu_r$$

$$\text{s.t.} \quad \sum_{q \in V'} a_q \lambda_q + \sum_{r \in D'} a_r \mu_r \geq b \quad (17)$$

$$\sum_{q \in V'} \lambda_q = 1 \quad (18)$$

$$\lambda_q \geq 0, \quad q \in V'; \quad \mu_r \geq 0, \quad r \in D'.$$

Siano  $\pi$  e  $\pi^0$  le variabili duali associate rispettivamente ai vincoli (17) e (18).

**Costo ridotto per punti estremi e direzioni estreme** Nel simplex standard, una colonna entra in base se il costo ridotto è negativo. Per i punti estremi, si verifica se esiste  $q \in V$  tale che:

$$c_q = c_q - [\pi^T \quad \pi^0] \begin{bmatrix} a_q \\ 1 \end{bmatrix} = c_q - \pi^T a_q - \pi^0 = c_q - \pi^T A v^q - \pi^0 < 0.$$

Per le direzioni estreme:

$$c_r = c_r - [\pi^T \quad \pi^0] \begin{bmatrix} a_r \\ 0 \end{bmatrix} = c_r - \pi^T a_r = c_r - \pi^T A d^r < 0.$$

**Auxiliary problem e pricing problem (PP)** In linea di principio si dovrebbe risolvere:

$$\min \left\{ \min_{q \in V} c_q, \min_{r \in D} c_r \right\},$$

ma non è affrontabile per enumerazione completa. Si riconduce allora a un **pricing problem**:

$$(PP) \quad \min (c^T - \pi^T A)x$$

$$\text{s.t.} \quad Dx \geq f$$

$$x \geq 0_n.$$

Si trascura il termine costante  $\pi^0$  e si usa lo stesso PP per trovare un vettore  $x \in \mathbb{R}^n$  che giochi il ruolo di punto estremo o direzione estrema. Sia  $z_{PP}^*$  il valore ottimo di PP.

### Casi per l'aggiunta di colonne

- Se  $z_{PP}^* = -\infty$  (PP illimitato inferiormente), si è trovata una **direzione estrema**  $x^*$  con costo ridotto negativo; si aggiunge a RMP una nuova variabile  $\mu_r$  associata alla colonna

$$\begin{bmatrix} Ax^* \\ 0 \end{bmatrix}.$$

- Se  $-\infty < z_{PP}^* < \pi^0$  (PP limitato e valore inferiore alla variabile duale del vincolo di convessità), si è trovato un **nuovo punto estremo**  $x^*$  con costo ridotto negativo; si aggiunge a RMP una nuova variabile  $\lambda_q$  associata alla colonna

$$\begin{bmatrix} Ax^* \\ 1 \end{bmatrix}.$$

- Se  $z_{PP}^* \geq \pi^0$ , non esiste alcun punto/direzione utile da aggiungere e ci si può fermare.

**Interpretazione intuitiva** In termini intuitivi, le variabili duali ottime ottenute risolvendo RMP suggeriscono una direzione

$$(c - A^T \pi)$$

lungo la quale si “sonda” il poliedro  $X$ . Se  $X$  è un politopo si può trovare solo un punto estremo; se  $X$  è illimitato si può trovare una direzione estrema. Non serve trovare tutti i punti e tutte le direzioni: interessano solo quelli con costo ridotto negativo, perché possono migliorare il valore ottenuto risolvendo RMP.

## 4 CAP 4 – Sistemi MRP/ERP e approccio JIT

**Recap** I metodi classici di gestione delle scorte risultano inadeguati in presenza di strutture di prodotto complesse, vincoli di capacità e ambienti non make-to-stock, poiché la propagazione dei fabbisogni lungo la distinta base può generare **amplificazione della variabilità**. In questo contesto si colloca l’evoluzione dalla logica MRP ai sistemi MRPII ed ERP, come risposta al problema del lot-sizing multilivello.

La logica **MRP** si fonda su un’ipotesi di capacità infinita, in cui il vincolo di capacità è surrogato tramite un **lead time fissato a priori**, e su una procedura ricorsiva di esplosione dei fabbisogni a partire dall’MPS. Gli ordini pianificati non sono esecutivi e il processo è soggetto al fenomeno del **nervosismo**, per cui piccole variazioni dell’MPS possono produrre grandi variazioni negli ordini, anche per effetto di bordo dovuto alla ripianificazione rolling horizon.

A livello di shop floor, la **Factory Physics** mette in relazione throughput, flow time e WIP tramite la **legge di Little** e mostra come la variabilità e l’elevata utilizzazione aumentino i tempi di attraversamento. L’approccio **Just-In-Time (Toyota)** mira a ridurre la variabilità alla fonte mediante produzione livellata, controllo **pull** del WIP e riduzione dei tempi di setup, evidenziando il legame strutturale tra variabilità, livelli di magazzino e prestazioni del sistema produttivo.

### 4.1 Limiti degli approcci classici di gestione delle scorte

**Inadeguatezza dei modelli tradizionali** I classici approcci di controllo delle scorte presentano **limiti severi** nei seguenti contesti:

- Ambienti *non make-to-stock*, come *make-to-order* e *assemble-to-order*, in cui viene ignorata la variabilità prevedibile;
- Presenza di **vincoli di capacità produttiva**, per cui tali modelli risultano più adatti a problemi retail;
- Strutture di prodotto **complesse**, rappresentate mediante distinte base multilivello.

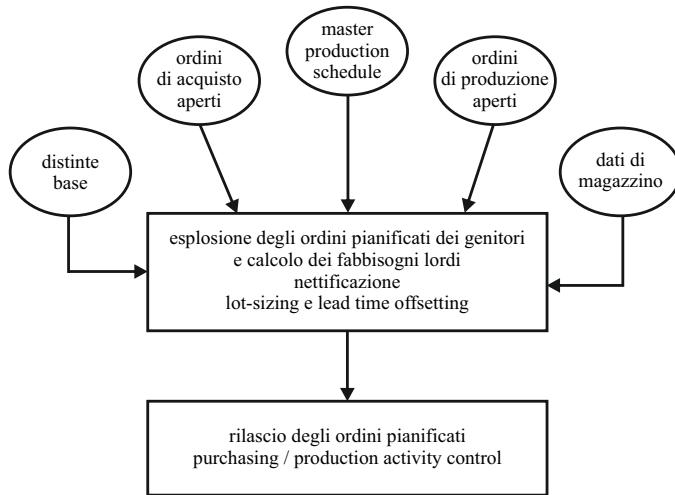
**Effetto di amplificazione della variabilità** Anche in presenza di una domanda regolare per il prodotto finito, la propagazione dei fabbisogni lungo la distinta base può indurre un’**amplificazione della variabilità** sui componenti a valle.

### 4.2 Evoluzione dai modelli MRP ai sistemi ERP

**Lot-sizing multilivello e difficoltà computazionali** In linea di principio è possibile costruire un modello MILP per il problema di **lot-sizing multilivello**, collegando domanda indipendente e domanda dipendente. Tali modelli risultano tuttavia difficili da risolvere, e erano computazionalmente impraticabili fino agli anni Settanta.

Negli anni Settanta sono stati introdotti i sistemi **MRP (Material Requirements Planning)**, basati su una **logica a capacità infinita**. Il vincolo di capacità produttiva viene rilassato, consentendo un **disaccoppiamento parziale tra item**, ma non tra domanda indipendente e dipendente.

L’interazione tra item viene anticipata e surrogata mediante un **lead time fissato a priori**.



**Evoluzione verso MRPII ed ERP** L’evoluzione successiva ha portato ai sistemi **MRPII (Manufacturing Resource Planning)**, che includono strumenti di verifica della capacità:

- **RCCP** (Rough Cut Capacity Planning).
- **CRP** (Capacity Requirement Planning): per verificare il soddisfacimento dei vincoli di capacità.

L’ulteriore evoluzione è rappresentata dai sistemi **ERP (Enterprise Resource Planning)**, caratterizzati dall’integrazione con le funzioni commerciali e finanziarie.

### 4.3 Logica MRP

**Assunzione di capacità infinita** La logica MRP assume **capacità infinita**: il vincolo di capacità produttiva non viene considerato esplicitamente, ma è rappresentato indirettamente dal lead time. Il lead time impone di lanciare gli ordini di produzione o di acquisto con sufficiente anticipo rispetto alla domanda. Il meccanismo di **lead time offsetting** consente di anticipare gli ordini pianificati rispetto ai fabbisogni.

**Record MRP** Ogni codice è descritto mediante un record MRP articolato per periodo, contenente: fabbisogni lordi, consegne attese, magazzino disponibile, fabbisogni netti e ordini pianificati. Prima del lead time offsetting è necessario calcolare i **fabbisogni netti**, ottenuti nettificando i fabbisogni lordi tenendo conto di: giacenze disponibili (*on-hand*) e ordini già emessi (*on-order*).

### 4.4 Esplosione dei fabbisogni e struttura del processo MRP

**Domanda indipendente e MPS** Per i prodotti finiti (*end item*), che costituiscono la radice della distinta base, i fabbisogni lordi sono definiti dal **Master Production Schedule (MPS)**. La logica MRP procede in modo **ricorsivo** a partire dagli end item, scendendo lungo le distinte base ed esplodendo i fabbisogni di ciascun codice fornendo i tempi di lancio degli ordini a tutti i livelli.

**Ordini pianificati, rilascio e regole di lot-sizing** I lotti corrispondono ai fabbisogni netti secondo la regola **lot-for-lot**. È importante comprendere la dinamica di un sistema MRP: gli ordini pianificati **non sono ordini esecutivi**; gli ordini pianificati dell’*action bucket*, una volta rilasciati, vengono trasformati in consegne attese. Un’ulteriore differenza tra ordini pianificati e ordini operativi è che, quando un ordine di

produzione viene rilasciato, vengono modificati i record relativi alle giacenze di magazzino dei componenti; una parte della giacenza viene allocata per l'ordine ed è da considerarsi non disponibile nel calcolo dei fabbisogni netti. I pacchetti MRP permettono di specificare un orizzonte temporale di release; solo gli ordini che ricadono all'interno di tale orizzonte dovrebbero essere rilasciati, in quanto gli altri sono soggetti a incertezze eccessive. Nel calcolo dei fabbisogni è possibile tenere conto di scorte di sicurezza; in questo caso, i fabbisogni netti vengono generati non quando il magazzino disponibile va sotto zero, ma quando va sotto un certo livello di soglia. Esiste una gamma di regole di lot-sizing, a quantità fissa o variabile, oppure euristiche per la minimizzazione dei costi totali di giacenza e di ordinazione.

## 4.5 Il problema del nervosismo

Il dimensionamento dei lotti a partire dai livelli alti della distinta base può generare effetti non intuitivi in presenza di variazioni del MPS. Le regole a quantità variabile sono soggette al fenomeno del **nervosismo** (*nervousness*).

Il nervosismo si manifesta quando **piccole variazioni nei fabbisogni** producono **grandi variazioni negli ordini pianificati**, anche a livelli inferiori della distinta base. Il fenomeno può portare a: ordini urgenti, instabilità del piano e risultati anti-intuitivi.

**Effetto di bordo, nervosismo e strategie di mitigazione** Un altro punto da considerare è l'**effetto di bordo** dovuto alla ripianificazione *rolling horizon*, in cui si aggiunge un *time bucket* all'orizzonte di pianificazione. L'aggiunta del fabbisogno relativo a tale time bucket può alterare l'accorpamento dei fabbisogni, con esiti imprevedibili. Esistono diversi modi per evitare il fenomeno del nervosismo. L'adozione di regole a quantità fissa permette di filtrare naturalmente variazioni di piccola entità, al costo di un aumento nel livello delle scorte. È anche possibile adottare strategie di gestione differenziata dell'orizzonte temporale di pianificazione, dette strategie di **time fencing**. Nell'immediato non è possibile cambiare nulla nell'MPS; in un periodo intermedio è possibile alterare l'MPS solo dopo specifica analisi e autorizzazione; è invece possibile cambiare l'MPS a piacere nei periodi più lontani. Infine, uno strumento utile per evitare il nervosismo e per risolvere situazioni anomale è l'uso dei **firm planned orders**; si tratta di ordini che non possono essere modificati dall'MRP quando le condizioni cambiano, ma soltanto dietro istruzione del pianificatore.

## 4.6 Master Production Scheduling e pianificazione della capacità

**Master Production Scheduling (MPS)** Il **Master Production Schedule (MPS)** costituisce l'input primario per la logica MRP ed è basato in parte su ordini cliente e in parte su attività di *forecasting* (demand planning). Nella logica MRP tradizionale, l'MPS può essere validato mediante moduli **RCCP (Rough Cut Capacity Planning)**. Non è detto che l'MPS venga costruito esclusivamente per i codici alla radice delle distinte base: può esistere domanda indipendente per parti di ricambio e, in contesti *assemble-to-order* (ATO), può essere preferibile adottare una struttura a due livelli, con un MPS a livello di moduli e un *Final Assembly Scheduling* a livello superiore.

**Capacity Requirements Planning (CRP)** La logica MRP è basata su un'assunzione di **capacità infinita**. È tuttavia possibile effettuare una verifica a posteriori del carico di lavoro rispetto alla capacità effettivamente disponibile tramite moduli di **Capacity Requirements Planning (CRP)**. La risoluzione manuale di eventuali situazioni di non ammissibilità risulta complessa, a meno di ricorrere a modelli di ottimizzazione o a procedure euristiche a capacità finita. Inoltre, per evitare ritardi, si tende spesso a gonfiare il lead time presunto, generando work in process che a sua volta allunga ulteriormente i lead time, dando luogo a un potenziale **circolo vizioso**.

## 4.7 Factory Physics e legge di Little

A livello di *shop floor*, le misure di prestazione fondamentali sono il **throughput**, il **flow time** (strettamente legato al lead time e comprensivo di tempi di attesa, lavorazione e movimentazione) e il **work in process (WIP)**, associato ai materiali in coda. Idealmente, si desidererebbe un throughput elevato con flow time e WIP contenuti; tuttavia, il raggiungimento simultaneo di tali obiettivi è fortemente influenzato dalla variabilità, sia prevedibile sia imprevedibile.

**Legge di Little** Un risultato fondamentale della teoria delle code è la **legge di Little**, che esprime la relazione tra WIP, throughput e flow time:

$$\text{WIP} = \text{throughput} \times \text{flow time}$$

La legge mette in evidenza il legame strutturale tra il livello di materiali in lavorazione e il tempo di attraversamento del sistema.

## 4.8 Ruolo della variabilità nel flusso dei materiali

**Modello di una singola macchina** Si consideri una singola macchina dotata di un buffer per il WIP. Si introducono le seguenti grandezze: il tempo medio di attesa in coda  $W_q$ , il tempo medio di lavorazione  $t_s$ , il tasso medio di servizio  $\mu = 1/t_s$ , il tasso medio di arrivo  $\lambda$ , che in condizioni di equilibrio coincide con il throughput, e la lunghezza media della coda  $L$ , che rappresenta il WIP. In tale contesto, la legge di Little può essere riscritta come:

$$L = \lambda (W_q + t_s)$$

L'utilizzazione del sistema è definita come  $u = \lambda/\mu$ , ed è compresa tra 0 e 1.

**Effetto della variabilità** Una coda del tipo  $G/G/1$  non è trattabile analiticamente in forma chiusa; tuttavia, una formula approssimata, esatta nel caso  $M/M/1$ , fornisce una stima del tempo medio di attesa in coda:

$$W_q \approx \left( \frac{C_a^2 + C_s^2}{2} \right) \left( \frac{u}{1-u} \right) t_s$$

dove  $C_a$  e  $C_s$  sono i coefficienti di variazione dei tempi di interarrivo e di servizio, rispettivamente. La relazione mostra come l'attesa in coda cresca rapidamente all'aumentare dell'utilizzazione e della variabilità.

## 4.9 Buffering Law e produzione livellata

**Buffering Law** Per ovviare agli effetti della variabilità è necessario introdurre dei **buffer**, che possono assumere la forma di magazzino o WIP, capacità in eccesso oppure tempo, sotto forma di lead time gonfiato. In assenza di una riduzione della variabilità, il sistema paga un prezzo in termini di WIP elevato, capacità sottoutilizzata e peggioramento del livello di servizio al cliente, manifestato da vendite perse, lead time lunghi e consegne in ritardo.

**Produzione livellata e fonti di variabilità** Uno dei fondamenti dell'approccio Toyota tradizionale è la **produzione livellata** (*production smoothing*). La variabilità non è legata esclusivamente a eventi casuali, come i guasti alle macchine, ma può derivare anche da batching dovuto ai tempi di setup, batching nella movimentazione (*wait to move*), scarso coordinamento nell'assemblaggio (*wait to match*) e variabilità della domanda riflessa nell'MPS.

## 4.10 Approccio Just-In-Time (Toyota)

**Principi dell'approccio Just-In-Time** L'approccio **Just-In-Time (JIT)**, sviluppato nell'ambito del sistema Toyota, si basa sull'idea di ridurre la variabilità mediante la ripetizione di un **mix di produzione ripetitivo** (*mixed-model*), realizzando una produzione livellata (*production smoothing*). Il presupposto fondamentale è la riduzione o l'annullamento dei tempi di setup, che consente di produrre lotti piccoli e frequenti. Il controllo del **work in process** avviene tramite un approccio **pull**, basato sul prelievo fisico dei materiali, in contrapposizione alla logica **push**, basata sul rilascio di ordini pianificati a partire da previsioni di fabbisogno. Lo strumento chiave del controllo pull è il **sistema kanban**, con possibile alternativa nel sistema *CONWIP*.

## 4.11 Toyota Goal Chasing

**Sequenziamento e regolarità dei flussi** A parità di mix produttivo, esistono diverse sequenze di assemblaggio che realizzano lo stesso numero di prodotti finiti. Il problema del **Toyota Goal Chasing** consiste nello scegliere una sequenza che renda il più possibile **regolare il fabbisogno di componenti** sulle linee laterali che alimentano la linea principale di assemblaggio. L'obiettivo è consentire il controllo delle linee a monte mediante un sistema pull, che richiede flussi regolari e prevedibili.

**Formalizzazione del problema** Si consideri una linea di assemblaggio che realizza  $N$  prodotti finiti sulla base di  $M$  moduli prodotti su linee laterali, con una distinta base piatta. Sia  $b_{ij}$  il numero di componenti di tipo  $j$  richiesti per assemblare un'unità del prodotto finito  $i$ , e sia  $Q_i$  il numero di finiti di tipo  $i$  previsti nel mix ripetitivo. Il fabbisogno per ciclo dei componenti di tipo  $j$  è dato da:

$$N_j = \sum_{i=1}^N b_{ij} Q_i$$

Indicando con  $Q = \sum_{i=1}^N Q_i$  il numero totale di assemblaggi per ciclo, il consumo cumulato ideale dei componenti di tipo  $j$  al passo  $k$  dovrebbe essere pari a:

$$\frac{kN_j}{Q}$$

**Funzione obiettivo del goal chasing** Sia  $X_{jk}$  il consumo cumulato effettivo del componente  $j$  al passo  $k$ . L'obiettivo del problema di goal chasing è minimizzare la distanza tra consumo ideale e consumo reale, misurata tramite la funzione:

$$\sum_{k=1}^Q \sum_{j=1}^M \left( \frac{kN_j}{Q} - X_{jk} \right)^2$$

**Ruolo dei tempi di setup nel JIT** Nel JIT viene posta forte enfasi sulla riduzione dei **tempi di setup**, poiché essa consente di ridurre la dimensione dei lotti e, di conseguenza, il livello medio delle giacenze. Tuttavia, l'impatto dei tempi di setup deve essere analizzato congiuntamente agli altri fattori del sistema produttivo. In particolare, la riduzione dei livelli di magazzino non dipende esclusivamente dalla diminuzione dei setup, ma anche dal tasso di produzione e dal grado di saturazione del sistema.

## 4.12 Modello di rotazione ciclica dei prodotti

**Impostazione del modello** Si consideri una linea su cui ruotano ciclicamente  $N$  prodotti. Per ciascun prodotto  $i$  si introducono il tasso di produzione  $p_i$  (pezzi per unità di tempo), il tasso di domanda  $d_i < p_i$ , assunto costante, e il tempo di setup  $s_i$ , assunto indipendente dalla sequenza. Il periodo di rotazione  $T_c$  deve essere ridotto al fine di contenere il livello medio di giacenza.

Indicando con  $T_i$  la durata del lotto del prodotto  $i$  in ciascuna rotazione, in condizioni di equilibrio deve valere:

$$p_i T_i = d_i T_c \quad \Rightarrow \quad T_i = \frac{d_i}{p_i} T_c$$

Inoltre, il periodo di rotazione deve soddisfare il vincolo:

$$T_c \geq \sum_{i=1}^N s_i + \sum_{i=1}^N T_i$$

L'eventuale *slack* è utile per assorbire guasti, ritardi ed effettuare attività di manutenzione.

**Limite inferiore del periodo di rotazione** Sostituendo l'espressione di  $T_i$  nel vincolo precedente, si ottiene il limite inferiore:

$$T_c \geq \frac{\sum_{i=1}^N s_i}{1 - \sum_{i=1}^N \frac{d_i}{p_i}}$$

Il risultato mostra che, oltre ai tempi di setup, il periodo di rotazione è fortemente influenzato dal rapporto tra tassi di domanda e tassi di produzione, evidenziando un legame diretto con i fenomeni di congestione già osservati nei modelli di coda.

## 5 CAP 5 – Schedulazione nella produzione e nei servizi

**Recap** La **schedulazione** consiste nell'assegnare **risorse** a **job** nel tempo, rispettando **vincoli tecnologici** e di **capacità** e ottimizzando una **misura di prestazione**; rispetto alla **pianificazione della produzione** (es., MRP) opera a un livello di dettaglio superiore. I job sono descritti da **operazioni** con **precedenze**, **tempi di processo**, **tempi di rilascio** e **due date**; tipicamente si assume lavorazione su **una macchina per volta**, **una sola operazione per macchina** e **assenza di interruzioni**, pur con possibili violazioni (lot streaming, batch, pre-emption).

Le strutture di problema derivano da diverse **precedenze** e layout di macchine (**macchina singola**, **parallele**  $P/Q/R$ , **flow shop**, **job shop**, **open shop**), e le prestazioni si misurano tramite funzioni di penalità sui tempi di completamento ( $C_i, F_i, L_i, T_i, E_i, U_i$ ), anche in forma **aggregata** (min-sum o min-max, con **pesi**  $w_i$ ). Alcune misure risultano **equivalenti** fino a costanti (lateness totale e flow time totale) e  $L_{\max}$  implica ottimalità anche per  $T_{\max}$ , con  $T_{\max} = \max\{L_{\max}, 0\}$ . Le misure **regolari** (non decrescenti in  $C_i$ ) permettono di ricavare il **timing** dal **sequenziamento** tramite **soluzioni attive**; con misure **non regolari** (es., somma pesata di earliness e tardiness) sequencing e timing non coincidono e la soluzione ottima non è necessariamente attiva o semiattiva.

La notazione di **Graham**  $\alpha|\beta|\gamma$  classifica i problemi tramite layout, vincoli aggiuntivi e misura obiettivo. Esistono casi risolvibili in modo **polinomiale** (regole **EDD**, **WSPT**, algoritmo di **Johnson**), ma in generale i problemi sono **NP-hard** e si ricorre a **euristiche** e **metaeuristiche**, includendo regole avanzate come **ATC**. La formulazione tramite **grafo disgiuntivo** rappresenta precedenze (archi congiuntivi) e vincoli di capacità (archi disgiuntivi), e il **cammino critico** determina il **makespan**; da tale grafo si può derivare un **MILP** per  $J//C_{\max}$ , utile come base per strategie *destroy and repair* pur non essendo trattabile in pratica. Nella ricerca locale, perturbare archi disgiuntivi sul **cammino critico** evita cicli ed è l'unica perturbazione utile per  $J//C_{\max}$ .

La procedura **shifting bottleneck** decomponne  $J//C_{\max}$  in problemi su macchina singola del tipo  $1/r_i/L_{\max}$ , usando **teste**  $h_{ib}$  (rilasci  $r_{ib}$ ) e **code**  $t_{ib}$  per costruire **due date locali**  $d_{ib} = K - t_{ib}$ ; il collo di bottiglia è la macchina con  $L_{\max}$  peggiore e viene congelata iterativamente fino a schedulare tutte le macchine, con possibili raffinamenti ed estensioni a  $J/r_i/L_{\max}$ .

## 5.1 Modelli classici di machine scheduling

Risolvere un problema di **schedulazione** richiede di assegnare **risorse** a **job** da eseguire nel tempo, rispettando **vincoli tecnologici** e di **capacità**, in modo da ottimizzare una **misura di prestazione**.

Un problema di scheduling si pone a un livello di dettaglio superiore rispetto a un problema di **pianificazione della produzione** (ad esempio a livello **MRP**). Esiste una grande varietà di problemi di scheduling, ma si inizia con la definizione di un **problema minimale**.

**Caratteristiche dei job** I job sono caratterizzati da:

- Un **set di operazioni** da eseguire, legate da **vincoli di precedenza**;
- **Tempi di lavorazione** per l'esecuzione su ogni macchina;
- **Date di rilascio e date di consegna.**

Assunzioni comuni sono che un job possa essere in lavorazione su **al più una macchina per volta**, che ogni macchina possa lavorare **un job per volta**, e che le lavorazioni **non possano essere interrotte**. Ognuna di tali assunzioni può essere violata (**lot streaming, processori batch, pre-emption**).

**Soluzioni e diagrammi di Gantt** Una soluzione di un problema di scheduling è caratterizzata dalle **sequenze di lavorazione** sulle macchine. Una soluzione può essere visualizzata mediante un **diagramma di Gantt**, che rappresenta graficamente l'allocazione temporale dei job sulle macchine.

Dall'esempio si ricava che una stessa istanza può ammettere più soluzioni ammissibili, caratterizzate da diverse sequenze sulle macchine.

## 5.2 Tipi di flusso

**Strutture di precedenza** Esistono diverse **strutture di precedenza** tra le operazioni dei job, che caratterizzano la **tecnologia del ciclo di lavorazione**. Sono possibili **strutture lineari**, oppure **strutture arbitrarie** rappresentabili mediante **grafi**. Le precedenze si traducono in **vincoli** di un problema di ottimizzazione.

Il **flusso dei materiali** caratterizza diverse strutture in termini di macchine:

- **macchina singola**;
- **macchine parallele** (identiche:  $p_{im} = p_i$ , correlate:  $p_{im} = p_{ism}$ , scorrelate);
- **flow shop**;
- **job shop**;
- **open shop**.

Oltre a tali strutture prototipali, si osservano varianti sul tema, e sono possibili **strutture ibride** (ad esempio **flexible flow shop**).

## 5.3 Misure di prestazione

Si indica con  $C_i$  il **tempo di completamento** del job  $i \in [n]$ , ovvero il tempo di completamento della sua **ultima operazione**.

**Funzioni di penalità** Un tipico modo per costruire misure di prestazione è associare ad ogni job  $J_i$  un termine di **penalità**  $\gamma_i(C_i)$ . I termini tipicamente utilizzati sono:

- **tempo di completamento**:  $\gamma_i(C_i) = C_i$ ;
- **flow time**:  $\gamma_i(C_i) = F_i = C_i - r_i$ ;

- **lateness**:  $\gamma_i(C_i) = L_i = C_i - d_i$ ;
- **tardiness**:  $\gamma_i(C_i) = T_i = \max\{C_i - d_i, 0\}$ ;
- **earliness**:  $\gamma_i(C_i) = E_i = \max\{d_i - C_i, 0\}$ ;
- **indicatore di ritardo**:  $\gamma_i(C_i) = U_i = \begin{cases} 1 & \text{se } C_i > d_i, \\ 0 & \text{se } C_i \leq d_i. \end{cases}$

**Pesi e priorità dei job** È possibile associare **pesi**  $w_i$  ai job, allo scopo di esprimere una **priorità**. Si può per esempio valutare per ogni job  $J_i$  la **tardiness pesata**  $w_i T_i$ .

Sulla base di questi “**mattoncini**” si possono costruire misure di prestazione **aggregate**, ottenendo funzioni del tipo:

$$\min \sum_{i=1}^n \gamma_i(C_i), \quad \min \max_{i=1, \dots, n} \gamma_i(C_i).$$

**Principali misure di prestazione aggregate** Le principali misure di prestazione aggregate sono:

- **flow time totale**:  $\sum_i F_i$ ;
- **flow time totale pesato**:  $\sum_i w_i F_i$ ;
- **massima lateness**:  $L_{\max} = \max_i L_i$ ;
- **tardiness totale pesata**:  $\sum_i w_i T_i$ ;
- **makespan**:  $C_{\max} = \max_i C_i$ ;
- **numero di job in ritardo**:  $\sum_i U_i$ .

**Equivalenza tra misure di prestazione** Alcune misure sono tra di loro **equivalenti**, nel senso che una soluzione ottima rispetto a una di esse è ottima anche per l'altra. Ad esempio, la **lateness totale** è equivalente al **flow time totale**:

$$\sum_{i=1}^n L_i = \sum_{i=1}^n C_i - \sum_{i=1}^n d_i = \sum_{i=1}^n F_i + \sum_{i=1}^n (r_i - d_i).$$

Le due misure differiscono per una **costante**. Inoltre, una soluzione ottima rispetto a  $L_{\max}$  è ottima anche rispetto a  $T_{\max}$  (ma non è assicurato il viceversa):  $T_{\max} = \max\{L_{\max}, 0\}$ .

**Misure di prestazione regolari** Le misure di prestazione finora elencate sono **funzioni non decrescenti** dei tempi di completamento; si parla in questo caso di **misure di prestazione regolari**. Nel caso di misure regolari, le informazioni di **tempistica** possono essere ricavate dal **sequenziamento**, in quanto ogni operazione è eseguita **al più presto (soluzioni attive)**.

**Misure di prestazione non regolari** Nel caso di **misure non regolari**, ciò non è più vero e il problema risulta **più complesso**. Tali misure sono utili per tenere conto dell'**inopportunità di completare un job prima della due date**, ad esempio adottando come misura:

$$\sum_{i=1}^n (w_i^T T_i + w_i^T E_i).$$

In questo caso esiste una regione in cui la funzione **decresce** al crescere del tempo di completamento, e **sequencing e timing non coincidono** e la soluzione ottima non è necessariamente attiva o semiattiva.

## 5.4 Notazione di Graham

Per classificare i problemi di schedulazione si utilizza la codifica a tre campi  $\alpha|\beta|\gamma$  dovuta a **Graham**.

**Campo  $\alpha$ : layout delle macchine** Il campo  $\alpha$  descrive il **layout delle macchine**:

- 1: macchina singola;
- $P, Q, R$ : macchine parallele identiche, correlate e scorrelate;
- $F$ : flow shop (ad esempio  $F2$ );
- $J$ : job shop.

**Campo  $\beta$ : vincoli aggiuntivi** Il campo  $\beta$  descrive la presenza di eventuali **vincoli aggiuntivi**:

- **res**: risorse aggiuntive;
- **prec**: vincoli di precedenza tra job;
- $r_i$ : tempi di rilascio non nulli;
- $\bar{d}_i$ : deadline hard;
- $s_{ij}$ : setup dipendenti dalla sequenza;
- **perm**: flow shop a permutazione, la frequenza di job si mantiene su tutte le macchine;
- **no-wait**: assenza di buffer tra operazioni: ragioni tecnologiche impongono che le operazioni vengano eseguite consecutivamente senza interruzioni.

**Campo  $\gamma$ : misura di prestazione** Il campo  $\gamma$  descrive la **misura di prestazione** da ottimizzare.

## 5.5 Algoritmi di soluzione

Esistono alcuni casi semplici risolvibili mediante **algoritmi polinomiali**:

- $1//L_{\max}$ : regola **EDD** che ordina i job in ordine di **due date** crescente;
- $1//\sum_i w_i c_i$ : regola **WSPT** (weighthed shortest processing time) che ordina i job in ordine decrescente del rapporto  $w_i/p_i$ ;
- $F2//C_{\max}$ : algoritmo di **Johnson** (in questo caso la soluzione ottima applica la stessa sequenza sulle due macchine).

**Problemi NP-hard ed euristiche** In generale, i problemi di schedulazione sono **NP-hard** e vengono affrontati mediante **euristiche**, sia costruttive sia iterative (**metaeuristiche** basate su ricerca locale). Le **regole di priorità** costituiscono una strategia costruttiva di base.

**Regola ATC (Apparent Tardiness Cost)** Esistono anche regole non banali, come la **regola ATC (Apparent Tardiness Cost)**, proposta per problemi  $J//\sum_i w_i T_i$ . L'espressione della priorità del job  $J_i$  sulla macchina  $M_j$  al tempo  $t$  è:

$$\frac{w_i}{p_{ij}} \exp\left(-\left[\frac{d_i - t - p_{ij} - \sum_{q=j+1}^{m_i} (W_{iq} + p_{iq})}{k\bar{p}}\right]^+\right),$$

dove  $w_i$  è il peso del job  $J_i$  e  $p_{ij}$  è il suo tempo di processo sulla macchina  $M_j$ ; la notazione  $[x]^+$  è equivalente a  $\max\{0, x\}$ ;  $W_{iq}$  è una stima del tempo di attesa di  $J_i$  sulla macchina  $M_q$ ;  $m_i$  è il numero di macchine che il job deve ancora visitare;  $k$  è un parametro da selezionare;  $\bar{p}$  è il tempo di processo medio dei job in coda.

**Interpretazione del termine di slack** L'espressione

$$t^* = d_i - p_{ij} - \sum_{q=j+1}^{m_i} (W_{iq} + p_{iq})$$

può essere interpretata come il massimo tempo di inizio di  $J_i$  su  $M_j$ , ottenuto sottraendo un lead time stimato dalla due date  $d_i$ . Se  $t \leq t^*$  si può ritenere di essere in tempo, altrimenti si corre il rischio di arrivare in ritardo. Se si è in tempo, il termine tra parentesi quadre è positivo, e la priorità cresce esponenzialmente al crescere di  $t$ .

Quando si è in ritardo, il termine tra parentesi quadre è negativo, per cui l'argomento dell'esponenziale è zero. La priorità è quindi  $w_i/p_{ij}$ , che coincide con la regola WSPT.

**Relazione con la tardiness totale pesata** Se molti job sono in ritardo, la tardiness totale pesata diventa una funzione obiettivo quasi equivalente al tempo di completamento totale pesato:

$$\sum_i w_i T_i = \sum_i w_i \max\{C_i - d_i, 0\} \approx \sum_i w_i C_i - \sum_i w_i d_i.$$

Il problema di minimizzazione di  $\sum_i w_i C_i$  è risolto, su macchina singola, dalla regola WSPT.

**Lookahead e ricerca locale** La miopia delle regole di priorità può essere ridotta introducendo un certo grado di lookahead, ad esempio mediante strategie di *beam search*. Una vasta classe di algoritmi iterativi si basa su perturbazioni locali della soluzione corrente, che definiscono una struttura di vicinato.

I principali problemi sono:

- sfuggire a minimi locali (tabu search, algoritmi genetici, questi richiedono attenta codifica e decodifica delle soluzioni);
- esplorare grandi vicinati in maniera efficiente (LNS, Large Neighborhood Search, utile anche per misure non regolari);
- evitare la creazione di cicli.

Molte di tali questioni richiedono una formulazione mediante un grafo disgiuntivo.

## 5.6 Grafi disgiuntivi

In un **grafo disgiuntivo**, i nodi rappresentano le operazioni, con l'aggiunta di nodi dummy iniziali e finali. Gli **archi congiuntivi** rappresentano vincoli di precedenza tecnologici, associati ai cicli di ciascun job. Gli **archi disgiuntivi** vanno orientati e rappresentano i vincoli di capacità per ogni macchina (una clique per macchina). Il **cammino critico**, ossia il cammino di lunghezza massima dal nodo iniziale al nodo finale, fornisce il makespan.

## 5.7 Modello MILP per $J//C_{max}$

**Formulazione** Sulla base del grafo disgiuntivo possiamo costruire un modello MILP per il problema  $J//C_{max}$ . Sia  $N = \{0, 1, \dots, N-1, N\}$  l'insieme dei nodi, dove 0 e  $N$  sono nodi dummy. Sia  $P$  l'insieme degli archi congiuntivi e  $D$  l'insieme degli archi disgiuntivi. La variabile binaria  $x_{ij}$ , se posta pari a 1, indica che l'operazione  $i$  precede l'operazione  $j$ .

$$\begin{aligned}
\min \quad & C_N \\
\text{s.t.} \quad & C_j \geq C_i + p_j & \forall (i, j) \in P \\
& C_j \geq C_i + p_j - M(1 - x_{ij}) & \forall (i, j) \in D \\
& C_i \geq C_j + p_i - Mx_{ij} & \forall (i, j) \in D \\
& x_{ij} \in \{0, 1\} & \forall (i, j) \in D \\
& C_i \geq p_i & \forall i \in N.
\end{aligned}$$

Il modello può essere adattato ad altre misure di prestazione, ma non è trattabile in pratica a causa di bound deboli. Tuttavia costituisce una base utile per metaeuristiche e approcci LNS (*destroy and repair*).

**Il ruolo del cammino critico** In una strategia di ricerca locale, perturbazioni arbitrarie delle sequenze possono creare cicli. Nel problema  $J//C_{\max}$ , se si perturbano archi disgiuntivi appartenenti al cammino critico, non si possono creare cicli, e tali perturbazioni sono le sole utili.

## 5.8 Procedura Shifting Bottleneck

**Idea generale** La procedura *shifting bottleneck*, originariamente pensata per il problema  $J//C_{\max}$ , sfrutta il grafo disgiuntivo per decomporre il problema in una sequenza di problemi del tipo  $1/r_i/L_{\max}$ , per i quali sono disponibili algoritmi branch-and-bound efficienti. Se una macchina  $M_b$  è il collo di bottiglia, le altre macchine possono essere trattate come risorse a capacità infinita, eliminando i relativi archi disgiuntivi corrispondenti alle altre macchine e mantenere solo quelle corrispondenti alle altre macchine.

**Teste e code nel grafo disgiuntivo** Si consideri il nodo  $O_{ib}$  del grafo, corrispondente all'operazione del job  $J_i$  sulla macchina  $M_b$ , e sia  $C_{ib}$  il suo tempo di completamento. Si consideri il **cammino critico** dal nodo fittizio iniziale al nodo  $O_{ib}$ . La sua lunghezza  $h_{ib}$  è la somma dei tempi di processo delle operazioni di  $J_i$  che precedono l'operazione  $O_{ib}$ , e ne costituisce la **testa**, ovvero il **tempo di rilascio** dell'operazione  $O_{ib}$ . In modo analogo si definisce la **coda**  $t_{ib}$  come la lunghezza del cammino critico dal nodo  $O_{ib}$  al nodo fittizio finale. In questo caso  $t_{ib}$  è la somma dei tempi di processo delle operazioni che seguono  $O_{ib}$ .

**Approssimazione del makespan** Si definisce una data di consegna per il job  $J_i$  sulla macchina  $M_b$ , approssimando il makespan come  $C_{\max} = \max_i\{C_i\} \approx \max_i\{C_{ib} + t_{ib}\}$ . Se si sottrae una costante arbitraria  $K$  dalla misura di prestazione, la soluzione ottima del problema non cambia, ottenendo il problema equivalente:

$$\max_i\{C_{ib} + t_{ib}\} - K = \max_i\{C_{ib} - (K - t_{ib})\} = \max_i\{C_{ib} - d_{ib}\},$$

dove è stata introdotta una **due date locale**  $d_{ib} = K - t_{ib}$ . Più lunga è la coda  $t_{ib}$ , più stretta è la data di consegna  $d_{ib}$ .

**Riduzione a  $1/r_i/L_{\max}$**  Il problema  $J//C_{\max}$  si riconduce quindi a un problema  $1/r_i/L_{\max}$ , in cui:  $r_{ib} = h_{ib}$ ,  $d_{ib} = K - t_{ib}$ . I tempi di rilascio sono dati dalle **teste**, mentre le date di consegna dipendono dalle **code**.

**Identificazione del collo di bottiglia** Un modo ragionevole per identificare il **collo di bottiglia** consiste nel risolvere i problemi  $1/r_i/L_{\max}$  per tutte le macchine, definendo per ciascuna le teste e le code delle operazioni. Il collo di bottiglia è la macchina che presenta il valore di  $L_{\max}$  peggiore. Dopo avere risolto i problemi su macchina singola, si individua il collo di bottiglia  $M_b$  e la relativa sequenza. È quindi possibile orientare gli archi disgiuntivi corrispondenti a  $M_b$  e rischedulare le altre macchine. In pratica, si risolvono i problemi su macchina singola per le rimanenti  $M - 1$  macchine, aggiornando teste

e code. Il procedimento prosegue fino ad avere schedulato tutte le macchine, congelandole una per volta. Sono possibili raffinamenti iterativi, e la strategia si estende facilmente al problema  $J/r_i/L_{max}$ ; sono state proposte anche ulteriori estensioni.

## 6 CAP 6 – Fondamenti microeconomici del pricing e modelli di scelta discreta

### 6.1 Driver del profitto

**Modello base del profitto** Il profitto è determinato da quattro fattori fondamentali ed è descritto dal seguente modello semplice:

$$\text{Profitto} = (\text{Prezzo} - \text{Costo variabile}) \cdot \text{Volume} - \text{Costo fisso}.$$

**Caso base** Si consideri il seguente caso:

- Prezzo: 100 euro per unità;
- Volume: 1 milione di unità;
- Costo variabile: 60 euro per unità;
- Costo fisso: 30 milioni di euro.

Il profitto risulta:

$$(100 - 60) \cdot 10^6 - 30 \cdot 10^6 = 10 \text{ milioni di euro.}$$

**Sensibilità del profitto ai fattori** Si considerano variazioni percentuali dei fattori nell'intervallo da  $-10\%$  a  $+10\%$  e il loro impatto percentuale sul profitto. Un aumento del prezzo del  $10\%$  comporta un raddoppio del profitto. Al contrario, una riduzione del costo variabile del  $10\%$  implica un aumento del profitto pari al  $50\%$ . Gli altri due fattori hanno un impatto minore. Il grafico risulta simmetrico rispetto alle variazioni positive e negative.

**Osservazioni** È importante notare che la modifica del prezzo è essenzialmente istantanea e priva di costo, a differenza degli investimenti in pubblicità o riduzione dei costi. Tuttavia, il modello è poco realistico poiché il prezzo influisce sul volume e il costo non è necessariamente una funzione affine del volume. È quindi necessario considerare le interazioni tra i fattori, pur utilizzando questo modello grezzo per sviluppare intuizioni.

### 6.2 Prezzo e volume

**Riduzione del prezzo** Si assuma una riduzione del prezzo del  $20\%$  rispetto al caso base. Per mantenere il profitto invariato, si deve risolvere:

$$10 = (80 - 60)V_{\text{new}} - 30 \Rightarrow V_{\text{new}} = 20 \text{ milioni.}$$

È quindi necessario un aumento del volume del  $100\%$ .

**Aumento del prezzo** Viceversa, se il prezzo aumenta del  $20\%$ , si ottiene:

$$10 = (120 - 60)V_{\text{new}} - 30 \Rightarrow V_{\text{new}} = \frac{2}{3} \cdot 20 \text{ milioni.}$$

È quindi possibile perdere un terzo del volume.

**Catene causali** È necessario modellare il legame tra prezzo e volume (domanda) ed essere consapevoli delle possibili catene causali:

- Prezzo → Volume → Ricavi → Profitto;
- Prezzo → Volume → Costi → Profitto;
- Prezzo → Prezzi dei concorrenti → Quota di mercato → Volume → Ricavi → Profitto;
- Prezzo del produttore → Prezzo del rivenditore → Volume → Ricavi → Profitto.

### 6.3 Pricing e revenue management

**Profitto e costi** L'obiettivo di un'impresa viene comunemente identificato nella massimizzazione del profitto. Tuttavia, l'analisi precedente mostra che ciò non si riduce alla sola minimizzazione dei costi.

**Revenue management** Se il profitto è ricavo meno costo, si aprono ulteriori possibilità. Le strategie di revenue e yield management sono diventate comuni a partire dall'industria aerea.

**Strategie di pricing** Le principali strategie di pricing includono:

- **Pricing basato sui costi:** valutazione del costo del prodotto e applicazione di un markup, ignorando le reazioni di consumatori e concorrenti;
- **Pricing basato sulla domanda:** il prezzo emerge dall'equilibrio tra domanda e offerta, modellabile in modo semplice in condizioni di monopolio;
- **Competizione:** modellata tramite la teoria dei giochi, sia tra imprese sia nell'interazione impresa–cliente.

### 6.4 Forme di competizione

**Tipologie di interazione** Le forme di competizione e interazione sono molteplici:

- competizione tra imprese (oligopolio vs. monopolio);
- interazioni lungo la supply chain (produttori e rivenditori);
- comportamento strategico dei clienti.

**Esempio: trasporto aereo** Un vettore aereo affronta:

- competizione diretta con altre compagnie;
- competizione con altri mezzi di trasporto (aereo vs. treno);
- competizione indiretta (viaggiare vs. meeting online).

### 6.5 Pricing nella pratica

**Cambio di paradigma** È necessario un cambio di paradigma:

$$\text{Design-Build-Price} \longrightarrow \text{Price-Design-Build}.$$

**Posizionamento di prezzo** Occorre selezionare una strategia di posizionamento coerente: luxury, premium, medium, low, ultra-low.

**Attributi del prodotto** La strategia di pricing dipende dagli attributi chiave del prodotto o servizio: funzionali, emotivi, simbolici, etici.

**Scelta del modello** È necessario scegliere il modello appropriato in funzione del contesto:

- acquisti ripetuti vs. acquisti singoli;
- beni di prima necessità vs. beni voluttuari;
- business-to-business (B2B) vs. business-to-consumer (B2C).

## 6.6 Meccanismi alternativi

**Oltre il prezzo** Il prezzo non è l'unico strumento per migliorare profitto e ricavi. Esistono diversi meccanismi alternativi:

- bundling e tying, sconti di quantità, pacchetti e servizi ancillari; talvolta unbundling;
- utilizzo di coupon;
- abbonamenti personalizzati;
- gestione della disponibilità di classi differenti;
- overbooking per gestire no-show e capacità non stoccabile;
- condivisione del rischio e progettazione di contratti e incentivi.

## 6.7 Parte 1 – Pricing e funzioni di domanda

**Un modello introduttivo: relazione lineare tra domanda e prezzo** La domanda è una variabile casuale, influenzata da una molteplicità di fattori, incluso (ma non solo) il prezzo. Come punto di partenza si introduce un modello semplice basato su una funzione di domanda lineare che lega prezzo e domanda:

$$d(p) = \alpha - \beta p. \quad (1)$$

**Interpretazione dei parametri** Un'ipotesi apparentemente ovvia è  $\beta > 0$ , ma esistono controesempi forniti dai beni di lusso e dai casi in cui il prezzo funge da segnale di qualità. L'intercetta  $\alpha$  rappresenta la domanda quando  $p = 0$ , un livello di prezzo per il quale il modello ha poco senso. In altri modelli, la domanda tende all'infinito per  $p \rightarrow 0$ , ma se la popolazione dei consumatori è finita,  $\alpha$  può essere interpretato come la dimensione della popolazione.

**Dominio di validità** Poiché la domanda non può essere negativa, il modello ha senso nell'intervallo di prezzo

$$p \in [0, p_{\lim}],$$

dove si definisce il prezzo limite

$$p_{\lim} := \frac{\alpha}{\beta}$$

come il prezzo massimo per il quale la domanda è positiva.

**Forma troncata della funzione di domanda** Il modello può essere riscritto in una forma potenzialmente migliore come

$$d(p) = (\alpha - \beta p)^+,$$

dove  $(x)^+ \equiv \max\{0, x\}$ .

**Funzione di domanda inversa** In microeconomia è comune invertire la funzione di domanda per ottenere il prezzo di mercato quando una quantità  $q$  viene prodotta e offerta sul mercato. Invertendo l'equazione (1), si ottiene la funzione di domanda inversa:

$$p(q) = a - bq,$$

dove  $a = \alpha/\beta$  e  $b = 1/\beta$ . Questa forma è utile nello studio della competizione basata sulle quantità nella microeconomia elementare, ma risulta meno naturale nel contesto di una modellazione statistica appropriata per il pricing management. Tuttavia, esistono mercati in cui il prezzo emerge da meccanismi complessi (ad esempio aste) che dipendono dalla disponibilità totale.

## 6.8 Massimizzazione dei ricavi e del profitto

**Profitto come funzione della quantità** Un obiettivo naturale è la massimizzazione del profitto. Nella sua forma più semplice, il profitto è dato da ricavi meno costi, e può essere espresso come funzione del prezzo o della quantità. La seconda scelta è più naturale per esprimere i costi:

$$\pi(q) = p(q) \cdot q - c(q) = r(q) - c(q),$$

dove la funzione di costo  $c(q)$  è una funzione generica della quantità e può tenere conto di costi fissi, economie o diseconomie di scala. Si introduce inoltre la funzione di ricavo

$$r(q) := p(q) \cdot q.$$

**Condizione di ottimalità** Assumendo che tutte le funzioni siano differenziabili e che la funzione di profitto complessiva sia concava, si applica la condizione di ottimalità del primo ordine:

$$\pi'(q^*) = r'(q^*) - c'(q^*) = 0 \Rightarrow r'(q^*) = c'(q^*).$$

La quantità ottimale è tale per cui il ricavo marginale e il costo marginale coincidono. Se il ricavo marginale è maggiore del costo marginale, il beneficio in termini di ricavi derivante da un aumento della produzione supera l'aumento dei costi, e conviene aumentare  $q$ . Se il costo marginale è maggiore del ricavo marginale, conviene ridurre  $q$ .

**Caso lineare con costo variabile costante** Nel caso di funzione di domanda inversa lineare e costo variabile costante, il profitto è una funzione quadratica concava:

$$\pi(q) = (a - bq) \cdot q - cq = (a - c) \cdot q - bq^2.$$

La condizione di ottimalità del primo ordine è:

$$\pi'(q) = (a - c) - 2bq,$$

da cui si ottiene:

$$q^* = \frac{a - c}{2b}.$$

**Osservazioni** La massimizzazione del profitto non è equivalente alla minimizzazione dei costi. La minimizzazione dei costi è appropriata quando una certa domanda deve comunque essere soddisfatta e occorre scegliere tra diverse tecnologie produttive o piani di produzione nel tempo. In alcuni settori il costo marginale di un'unità aggiuntiva è trascurabile (industria aerea, servizi web) o il costo di produzione è un costo sommerso (markdown management). In questi casi può essere sensata la massimizzazione dei ricavi, naturalmente espressa come funzione del prezzo:

$$\max r(p) = p \cdot d(p) = p \cdot (\alpha - \beta p) = \alpha p - \beta p^2.$$

**Soluzione ottima per la massimizzazione dei ricavi** La soluzione ottima è:

$$p^* = \frac{\alpha}{2\beta} \quad \Rightarrow \quad d^* = \frac{\alpha}{2} \quad \Rightarrow \quad r^* = \frac{\alpha^2}{4\beta}.$$

**Interpretazione geometrica** La soluzione può essere interpretata geometricamente: il prezzo ottimale è il punto medio dell'intervallo dei prezzi e massimizza l'area del rettangolo associato ai ricavi.

## 6.9 Stima del modello

**Problema della stima** Nella microeconomia elementare si utilizzano modelli deterministici, ma anche un modello di domanda lineare apparentemente banale come

$$d(p) = \alpha - \beta p$$

contiene parametri ignoti che devono essere stimati.

**Apprendimento e sperimentazione** Anche assumendo che il modello lineare sia corretto, esiste un problema rilevante: il costo della procedura di stima. È necessario apprendere online piuttosto che offline, e la pianificazione degli esperimenti è cruciale. Quali prezzi utilizzare per apprendere efficacemente i parametri? L'attenzione dovrebbe concentrarsi sulla pendenza  $\beta$ , poiché l'intercetta  $\alpha$  è principalmente un costrutto matematico.

**Errore di stima** L'errore standard della stima della pendenza è dato da:

$$SE_{\beta} = \sqrt{\frac{\sigma_{\varepsilon}^2}{\sum_{k=1}^n (p_k - \bar{p})^2}},$$

dove  $\sigma_{\varepsilon}$  è la deviazione standard (ignota) degli errori. Questo mostra che maggiore è la dispersione dei prezzi sperimentati, migliore è l'apprendimento. Tuttavia, prezzi molto bassi o molto elevati possono danneggiare i ricavi, generando un trade-off tra apprendimento e costo dell'apprendimento.

**Problemi pratici aggiuntivi** Nella pratica emergono ulteriori difficoltà:

- prezzi commerciali: non tutti i prezzi sono applicabili;
- grandi variazioni di prezzo enfatizzano la non linearità;
- ruolo del tempo: nei beni di moda il tempo è essenziale e il periodo di apprendimento è limitato;
- fattori confondenti: una variazione della domanda può dipendere da elementi diversi dal prezzo;
- problemi operativi ed esecutivi, ad esempio la modifica dei cartellini in un negozio fisico;
- sperimentazione eccessiva può generare insoddisfazione nei consumatori e indurre comportamenti strategici.

## 6.10 Parte 1 – Limiti dei modelli di domanda lineare

**Criticità del modello lineare** Il modello di domanda lineare presenta diversi limiti:

- l'ipotesi di pendenza costante (sensibilità al prezzo) non è realistica;
- si basa su un'ipotesi irrealistica sulla willingness-to-pay;
- si assume che la pendenza sia negativa; tuttavia, nel caso di beni di lusso e quando il prezzo funge da segnale di qualità, si possono osservare comportamenti localmente differenti;

- è un modello statico: l'introduzione del tempo è essenziale nelle strategie dinamiche di markdown;
- non si considera il ruolo della capacità (si veda, ad esempio, il peak-load pricing);
- si considera la vendita di un singolo bene o servizio (si veda tying, bundling e assortment management);
- non si considera la scelta del consumatore tra alternative che possono differire non solo per il prezzo;
- non si considerano aspetti comportamentali e asimmetrie informative; la psicologia del pricing può essere rilevante.

## 6.11 La funzione di costo

**Costo fisso e costo variabile** È comune considerare una componente di costo fisso e una variabile; la forma più semplice di costo variabile è lineare:

$$c(q) = F + cq.$$

**Costo fisso vs. costo con carica fissa** Una funzione di costo alternativa è:

$$c(q) = \begin{cases} F + cq & \text{se } q > 0, \\ 0 & \text{se } q = 0. \end{cases}$$

che può essere scritta come

$$c(q) = F \cdot \delta(q) + cq,$$

dove  $\delta(q) = 1$  se  $q > 0$  e  $\delta(q) = 0$  altrimenti. Per evitare ambiguità, si parla di **costo fisso** nel primo caso e di **carica fissa** nel secondo. Si noti che questa distinzione può dipendere dalla scala temporale e dal livello decisionale gerarchico: alla scala temporale appropriata, tutti i costi sono variabili (si veda il concetto di costi semivariabili).

**Costo marginale e costo medio** Data una funzione di costo, si definiscono i seguenti concetti:

- costo marginale  $c'(q)$ ;
- costo medio  $\frac{c(q)}{q}$ .

**Regolarità della funzione di costo** Per definire il costo marginale è necessario assumere la differenziabilità. Oltre alle cariche fisse, discontinuità possono essere introdotte da sconti di quantità all-units, mentre sconti incrementali introducono punti angolosi. Quando la produzione è discreta e  $q$  assume valori interi, il costo marginale può essere approssimato come

$$c(q+1) - c(q).$$

**Economie e diseconomie di scala** Il costo marginale può essere crescente o decrescente. Quando il costo marginale è decrescente, assumendo la differenziabilità, si ha  $c''(q) < 0$ , e quindi la funzione di costo è concava. Questo modella un'economia di scala, ossia un aumento dell'efficienza all'aumentare della produzione. Al contrario, una funzione di costo convessa, caratterizzata da un costo marginale crescente, modella una diseconomia di scala.

## 6.12 Funzioni di domanda

**Classi di modelli di domanda** Esiste una grande varietà di modelli di domanda:

- si può modellare la domanda aggregata dell'intero mercato, di un canale selezionato o di una parte del mercato, oppure la domanda di un singolo individuo;
- si possono considerare acquisti ripetuti oppure no, a seconda della natura del bene o servizio (beni stock vs. beni flusso) e dell'orizzonte temporale;
- si può considerare o meno l'incertezza.

**Input e output dei modelli** I fattori di input possono includere tempo, vendite passate, prezzi dei concorrenti o di beni alternativi. L'output del modello può essere:

- una variabile reale (domanda aggregata, eventualmente approssimante un valore intero elevato);
- una variabile intera (quantità acquistata da un singolo consumatore);
- una variabile binaria (acquisto o non acquisto);
- una variabile categoriale o un vettore (bundle o portafoglio di beni acquistati).

## 6.13 La funzione di domanda lineare

**Forma del modello** Per scopi principalmente illustrativi, si è già considerata la funzione di domanda lineare:

$$d(p) = (\alpha - \beta p)^+.$$

**Questioni aperte** È naturale interrogarsi sulla validità di un modello così semplice:

- come può essere giustificato dal punto di vista economico? È necessario indagare una fondazione microeconomica;
- è un modello validato empiricamente o è contraddetto dal buon senso e dal comportamento reale dei consumatori?

**Pendenza della domanda** Assumendo la differenziabilità, una caratteristica fondamentale di ogni funzione di domanda è la sua pendenza  $d'(p)$ , che per buon senso dovrebbe essere negativa, come nel modello lineare. In effetti è solitamente negativa, ma esistono eccezioni:

- segnale di qualità;
- beni di lusso e consumo vistoso.

Possono quindi esistere intervalli di prezzo per i quali la pendenza è positiva, un fatto confermato da evidenze empiriche (si vedano esempi in Simon, 2015).

## 6.14 Funzioni di domanda: elasticità puntuale

**Elasticità di prezzo della domanda** Poiché il valore della pendenza dipende dalla scala, ovvero dalle unità di misura della domanda, una misura più appropriata è l'elasticità di prezzo della domanda:

$$\frac{\Delta d/d}{\Delta p/p} = \frac{\Delta d}{\Delta p} \cdot \frac{p}{d}.$$

Facendo tendere  $\Delta p \rightarrow 0$ , si ottiene l'elasticità puntuale:

$$\varepsilon(p) = -\frac{d'(p)p}{d(p)}, \quad (2)$$

dove il segno meno è introdotto per comodità, poiché la pendenza è tipicamente negativa.

**Classificazione della domanda** Le funzioni di domanda possono essere classificate come:

- elastiche, quando  $\varepsilon(p) > 1$  (perfettamente elastiche se pari a 1);
- a elasticità unitaria, quando  $\varepsilon(p) = 1$ ;
- anelastiche, quando  $\varepsilon(p) < 1$  (perfettamente anelastiche se pari a 0).

La natura della domanda può dipendere dal punto specifico considerato.

## 6.15 Esempio: funzione di domanda lineare

**Elasticità vs. pendenza** L'elasticità non deve essere confusa con la pendenza, come mostra il seguente esempio. Il modello di domanda lineare presenta una pendenza costante, ma ciò non implica un'elasticità puntuale costante.

**Elasticità nel caso lineare** Per una funzione di domanda lineare, l'elasticità è:

$$\varepsilon(p) = \frac{\beta p}{\alpha - \beta p}.$$

Essa cresce da zero a +1 nell'intervallo  $[0, p_{\lim}]$ . L'elasticità è pari a 1 per

$$p = \frac{\alpha}{2\beta} = p^*,$$

ossia nel punto medio dell'intervallo dei prezzi. La domanda è anelastica per  $p < p^*$  ed elastica per  $p > p^*$ . Nel limite, quando  $p \rightarrow p_{\lim} = \alpha/\beta$ , diventa perfettamente elastica: una piccola riduzione di prezzo fa passare la domanda da  $d = 0$  a  $d > 0$ . L'elasticità non è costante per una funzione di domanda lineare.

**Domanda a elasticità costante** È naturale chiedersi se esista una funzione di domanda con elasticità costante. Imponendo

$$\frac{d'(p)p}{d(p)} = -\varepsilon,$$

si ottiene l'equazione differenziale ordinaria:

$$d'(p) = -\varepsilon \cdot \frac{d(p)}{p},$$

che ammette come soluzione:

$$d(p) = c \cdot p^{-\varepsilon},$$

dove  $c = d(1)$ .

**Nota tecnica** Per risolvere l'equazione differenziale precedente, si procede per separazione delle variabili:

$$\frac{dy}{y} + \varepsilon \frac{dx}{x} = 0,$$

dove  $y \equiv d(p)$  e  $x \equiv p$ . Integrando si ottiene:

$$\int \frac{dz}{z} + \varepsilon \int \frac{dz}{z} = 0 \Rightarrow \log y + \varepsilon \log x = K,$$

dove  $K$  è una costante di integrazione, da cui:

$$\log(y \cdot x^\varepsilon) = K \Rightarrow y = cx^{-\varepsilon},$$

ponendo  $c = e^K$ . Imponendo  $p = 1$ , si ottiene:

$$d(1) = c.$$

**Stima tramite trasformazione logaritmica** È possibile derivare un legame tra domanda a elasticità costante e domanda lineare mediante una trasformazione dei dati basata sui logaritmi:

$$d = cp^{-\varepsilon} \Rightarrow \log d = \log c - \varepsilon \log p.$$

In questo modo, gli strumenti standard di regressione lineare possono essere utilizzati per stimare il modello. L'uso dei logaritmi è utile per semplificare modelli basati su effetti moltiplicativi anziché additivi.

**Elasticità come derivata logaritmica** In termini di incrementi, per  $x > 0$  vale:

$$\frac{d \log x}{dx} = \frac{1}{x} \Rightarrow \Delta \log x \approx \frac{\Delta x}{x}.$$

Di conseguenza, l'elasticità può essere riscritta come:

$$\frac{\Delta d/d}{\Delta p/p} \approx \frac{d \log d(p)}{d \log p}.$$

**Ricavi con domanda a elasticità costante** Nel caso di domanda a elasticità costante, il ricavo è:

$$R(p) = p \cdot cp^{-\varepsilon} = cp^{1-\varepsilon}.$$

Il ricavo è costante nel caso di elasticità unitaria, crescente nel caso di domanda anelastica e decrescente nel caso di domanda elastica. Pertanto, una riduzione di prezzo aumenta i ricavi solo se la domanda è elastica.

## 6.16 Parte 1 – Interpretare i modelli di domanda: willingness-to-pay

**Confronto tra modelli e razionale economico** Come si possono confrontare diversi modelli di domanda? La funzione di domanda lineare è un modello sensato? Un modello lineare può essere (nel migliore dei casi) un'approssimazione locale, ma è necessario comprendere più a fondo la razionalità economica alla base di tale modello. Ciò significa che occorre capire che cosa determina realmente la domanda.

**Reserve price e willingness-to-pay** Un meccanismo di base dietro i modelli di domanda è il concetto di prezzo di riserva o willingness-to-pay: ogni consumatore è caratterizzato dal prezzo massimo al quale è disposto ad acquistare un bene.

**Esempio con segmentazione: studenti e non studenti** Come esempio semplice, si consideri la seguente tabella:

	Studenti	Non studenti
Willingness-to-pay	e 5	e 10
Dimensione del mercato	200	300

Questo esempio ipotetico descrive la willingness-to-pay per un certo tipo di spettacolo, se la popolazione può essere segmentata in due gruppi, studenti e non studenti, con diversa willingness-to-pay.

**Domanda a tratti** Questo produce una funzione di domanda a tratti costante, come indicato in figura.

**Popolazione differenziata e funzione a gradini** Si immagini ora una popolazione più differenziata di  $N$  consumatori potenziali, indicizzati da  $k = 1, \dots, N$ , ognuno con una willingness-to-pay unica:

$$p_k = p_{\max} \frac{k}{N}.$$

L'indice  $k = 1$  corrisponde al consumatore meno disposto a pagare, tale che tutti acquistano e la domanda è  $N$ . Per  $k = N$ , soltanto un consumatore acquista. La domanda è legata a  $k$  da:

$$d(p_k) = N - k + 1, \quad k = 1, \dots, N,$$

da cui risulta una funzione decrescente a gradini. L'intuizione suggerisce che, nel limite continuo, assumendo una willingness-to-pay uniformemente distribuita, si ottenga una funzione di domanda lineare. Funzioni di domanda alternative possono essere definite assumendo distribuzioni differenti del prezzo di riserva e studiando la sensibilità della domanda rispetto al prezzo.

## 6.17 Surplus del consumatore e discriminazione di prezzo

**Definizione di consumer surplus** Si consideri un consumatore disposto a pagare e100 per un bene. Se il prezzo è e70, si dice che il consumer surplus è la differenza tra willingness-to-pay e prezzo pagato, in questo caso e30. Quando viene applicato un prezzo unico, una parte di consumer surplus è ottenuta dall'insieme dei consumatori che acquistano.

**Consumer surplus in forma geometrica** Geometricamente, ciò è visualizzabile come il triangolo  $CS$  in figura, con area:

$$CS = \frac{1}{2} \left( \frac{\alpha}{\beta} - p_0 \right) (\alpha - \beta p_0) = \frac{1}{2\beta} (\alpha - \beta p_0)^2,$$

dove  $p_0$  è il prezzo scelto. Si noti che il consumer surplus è nullo quando  $p_0 = p_{\lim} = \alpha/\beta$ , ed è pari a  $\alpha^2/(2\beta)$ , ossia l'area totale del triangolo, quando  $p_0 = 0$ . In figura, il ricavo raccolto è rappresentato dall'area rettangolare  $REV$ .

**Interpretazione economica** In microeconomia, il consumer surplus è una misura approssimata dell'utilità del consumatore derivante dallo scambio. Guardando dalla prospettiva opposta, il consumer surplus è, in un certo senso, ricavo perso per l'impresa. Lo stesso vale per il triangolo  $LR$ , che rappresenta ricavo perso dovuto ai consumatori che non acquistano perché  $p_0$  eccede la loro willingness-to-pay.

**Prezzo unico e perdite di ricavo** Considerando solo i ricavi, l'obbligo di scegliere un solo prezzo genera perdite su entrambi i lati: consumer surplus (opportunità mancata di profitto maggiore) e vendite perse. Idealmente, l'impresa vorrebbe applicare un prezzo diverso a ciascun individuo, pari alla sua willingness-to-pay (assumendo costo marginale nullo).

**Discriminazione di prezzo** Naturalmente, fissare un prezzo individuale non è pratico, per non dire apertamente illegale. Come può un'impresa estrarre almeno parzialmente consumer surplus? La chiave è la discriminazione di prezzo basata sulla segmentazione dei consumatori, illustrata da un esempio semplice (tratto da Shy, 2008, p. 6).

## 6.18 Esempio di discriminazione di prezzo

**Caso senza discriminazione** Si consideri nuovamente la willingness-to-pay della tabella precedente. Se non è possibile discriminare, cioè si deve applicare lo stesso prezzo a tutta la popolazione, ci sono due scelte:

- Opzione 1: vendere a e10, con ricavo

$$10 \cdot 300 = \text{e}3000.$$

- Opzione 2: vendere a e5, con ricavo

$$5 \cdot (200 + 300) = \text{e}2500.$$

È quindi preferibile chiedere il prezzo più alto, escludendo gli studenti dall'acquisto del biglietto.

**Caso con discriminazione** Se si può discriminare e applicare prezzi diversi, il profitto aumenta:

$$5 \cdot 200 + 10 \cdot 300 = \text{e}4000.$$

Una tale politica è praticabile se si possono identificare gli studenti, ad esempio richiedendo un tesserino, e prevenire l'arbitraggio (uno studente che compra a e5 e rivende a un non studente a e10), che avrebbe un effetto di diluizione sul ricavo.

**Tipi di discriminazione** Applicando discriminazione di prezzo, l'impresa riesce a estrarre consumer surplus. D'altra parte, una sottopopolazione che sarebbe esclusa dallo scambio può partecipare. In generale, non è detto che si ottenga una soluzione win-win. Inoltre, potrebbe non essere possibile discriminare in modo netto. Esistono diversi tipi di discriminazione:

- discriminazione completa: ogni consumatore paga un prezzo specifico;
- segmentazione diretta: segmenti identificabili pagano prezzi diversi;
- segmentazione indiretta: si offrono varianti di prodotto/servizio e i consumatori scelgono; ciò funziona anche con caratteristiche non identificabili.

Esempi: acquisto anticipato e prenotazione; coupon di sconto.

## 6.19 Formalizzare willingness-to-pay e consumer surplus

**Distribuzione della willingness-to-pay** Dopo un'introduzione informale, si passa a un quadro più formale. La willingness-to-pay è il prezzo massimo (prezzo di riserva) che un consumatore è disposto a pagare per un bene. La distribuzione della willingness-to-pay nella popolazione determina la forma della funzione di domanda.

**Modello continuo e funzione  $w(x)$**  Si consideri un modello continuo e si definisca  $w(x)$  come distribuzione della willingness-to-pay. Essa svolge un ruolo simile alla densità di una variabile casuale, nel senso che

$$\int_{p_1}^{p_2} w(x) dx$$

è la frazione di popolazione con willingness-to-pay tra  $p_1$  e  $p_2$ . Quindi:

$$d(p) = D_{\max} \int_p^{+\infty} w(x) dx,$$

dove  $D_{\max} = d(0)$  è la domanda massima ottenibile (assumendo una popolazione finita). Poiché

$$d'(p) = -D_{\max} w(p),$$

si può legare domanda e willingness-to-pay tramite:

$$w(p) = -\frac{d'(p)}{D_{\max}}. \quad (3)$$

**Surplus del consumatore aggregato** Se un consumatore disposto a pagare  $p_1$  paga invece  $p_0 < p_1$ , gode di un consumer surplus  $p_1 - p_0$ . Dato  $w(x)$ , si integra il surplus al prezzo  $p_0$  sulla popolazione che acquista e si definisce il consumer surplus totale (netto):

$$S(p_0) = D_{\max} \int_{p_0}^{+\infty} w(x) (x - p_0) dx.$$

Usando (3) e integrando per parti:

$$\begin{aligned} S(p_0) &= - \int_{p_0}^{+\infty} d'(x) (x - p_0) dx = - \int_{p_0}^{+\infty} d'(x)x dx + p_0 \int_{p_0}^{+\infty} d'(x) dx \\ &= - \left[ d(x)x \Big|_{p_0}^{+\infty} - \int_{p_0}^{+\infty} d(x) dx \right] - p_0 d(p_0) = \int_{p_0}^{+\infty} d(x) dx. \end{aligned}$$

Quindi, il consumer surplus è l'area sotto la curva di domanda, a destra del prezzo applicato  $p_0$ .

## 6.20 Esempio: consumer surplus con domanda lineare e costo

**Dati e prezzo limite** Si consideri un articolo con costo unitario di produzione e3 e funzione di domanda aggregata:

$$d(p) = (1000 - 100p)^+.$$

Il prezzo limite che comporta domanda nulla è:

$$p_{\text{lim}} = \frac{1000}{100} = \text{e}10.$$

**Prezzo ottimo, domanda, profitto e surplus** Il prezzo ottimo è:

$$p^* = \frac{1000 + 100 \cdot 3}{2 \cdot 100} = 6.5,$$

con domanda, profitto e consumer surplus dati rispettivamente da:

$$d(p^*) = 1000 - 100 \cdot 6.5 = 350,$$

$$\pi^* = 350 \cdot (6.5 - 3) = \text{e}1225,$$

$$S(p^*) = \frac{(10 - 6.5) \cdot 350}{2} = 612.5.$$

Il consumer surplus è l'area ombreggiata ed è interpretabile come la perdita di opportunità per l'impresa dovuta all'impossibilità di chiedere il prezzo di riserva a ciascun consumatore.

**Obiettivo di estrarre consumer surplus** Un obiettivo naturale per l'impresa è “estrarre” consumer surplus almeno parzialmente tramite una forma di discriminazione di prezzo. Se l'impresa è in grado di discriminare, migliorerà il proprio profitto. L'intuizione suggerisce che ciò avvenga necessariamente a spese dei consumatori. L'esempio seguente mostra che non è necessariamente così, se consumatori con willingness-to-pay minore possono acquistare a un prezzo minore.

## 6.21 Esempio: consumer surplus e discriminazione di prezzo

**Domanda aggregata come somma di due popolazioni** Si consideri la stessa funzione di domanda dell'esempio precedente, ma si immagini che derivi dall'aggregazione di due popolazioni con funzioni di domanda:

$$d_1(p) = \min \{500, (1000 - 100p)^+\}, \quad d_2(p) = (500 - 100p)^+.$$

Ciò implica che la popolazione 1 include consumatori con willingness-to-pay maggiore di e5, mentre la popolazione 2 include consumatori con willingness-to-pay minore di e5.

**Prezzo ottimo per la popolazione 1** Se l'impresa può discriminare senza arbitraggio o diluizione, il prezzo ottimo per la popolazione 1 resta lo stesso di prima, così come domanda e profitto:

$$p_1^* = \text{e}6.5, \quad d_1(p_1^*) = 350, \quad \pi_1^* = \text{e}1225.$$

**Prezzo ottimo per la popolazione 2** L'impresa può vendere anche alla popolazione 2, con prezzo ottimo, domanda e profitto:

$$p_2^* = \frac{500 + 100 \cdot 3}{2 \cdot 100} = \text{e}4,$$

$$d_2(p_2^*) = 500 - 100 \cdot 4 = 100,$$

$$\pi_2^* = 100 \cdot (4 - 3) = \text{e}100.$$

**Profitto totale e surplus addizionale** Poiché è possibile vendere sopra costo alla popolazione 2 senza ridurre il ricavo dalla popolazione 1, l'impresa ottiene un profitto addizionale e il profitto totale è:

$$\pi^* = \pi_1^* + \pi_2^* = \text{e}1325.$$

L'impresa beneficia della segmentazione, ma anche i consumatori. Esiste un consumer surplus addizionale, poiché 100 consumatori possono acquistare a un prezzo inferiore al proprio prezzo di riserva:

$$S_2(p_2^*) = \frac{100 \cdot (5 - 4)}{2} = 50.$$

Il surplus totale è composto dai due triangoli ombreggiati.

## 6.22 La funzione logit

**Da willingness-to-pay a domanda** Integrare la densità di willingness-to-pay, analogamente a quanto si fa con una PDF per ottenere una CDF, fornisce una funzione di domanda. Un modello di domanda lineare non deriva da una distribuzione sensata della willingness-to-pay; è ragionevole ipotizzare una funzione “a campana” con un massimo (moda) concentrato attorno al prezzo di riserva più tipico.

**Probit e logit** Nella letteratura di statistica e marketing si utilizzano due funzioni per questo e altri scopi, che portano a modelli probit e logit. I modelli probit derivano dalla CDF di una normale standard, mentre i modelli logit si basano sulla funzione logistica (sigmoide).

**Definizione della funzione logistica** La funzione logistica è definita come:

$$L(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x},$$

e presenta la classica forma a S. Si osservi che  $L(0) = 0.5$  e che la funzione logit ha la proprietà di simmetria:

$$L(-x) = 1 - L(x).$$

**Derivata della funzione logistica** In figura si mostra anche la derivata:

$$L'(x) = \frac{e^x}{(1 + e^x)^2} = L(x)(1 - L(x)).$$

La derivata ha la proprietà di simmetria  $L'(x) = L'(-x)$  e il suo massimo si ha in  $x = 0$ .

**Funzione di domanda logit decrescente** Poiché le funzioni di domanda devono essere decrescenti, si deve ribaltare l'asse  $x$  da sinistra a destra e introdurre fattori di scala e traslazione sensati nella funzione logit, ottenendo la seguente funzione di domanda logit (decrescente):

$$d(p) = \frac{c e^{-(\alpha + \beta p)}}{1 + e^{-(\alpha + \beta p)}},$$

dove  $\beta, c > 0$ . Più grande è  $\beta$ , maggiore è la sensibilità al prezzo, e la funzione è più ripida per

$$\alpha + \beta p^* = 0 \Rightarrow p^* = -\frac{\alpha}{\beta},$$

il che suggerisce  $\alpha < 0$ . Questo è il valore per cui la densità di willingness-to-pay è massima e può essere considerato un “prezzo di mercato”.

**Willingness-to-pay corrispondente** La willingness-to-pay corrispondente è:

$$w(x) = -\frac{d'(p)}{d(0)} = \frac{K e^{-(\alpha+\beta x)}}{(1 + e^{-(\alpha+\beta x)})^2},$$

dove  $K = \beta c/d(0)$ .

**Esempio numerico** Si consideri il caso  $c = 1000$ ,  $p^* = 20$  e  $\beta = 0.5$ . In tal caso,

$$\alpha = -p^*\beta = -20 \cdot 0.5 = -10,$$

e la domanda per  $p = 0$  è:

$$d(0) = \frac{1000 \cdot e^{-\alpha}}{1 + e^{-\alpha}} = \frac{1000 \cdot e^{10}}{1 + e^{10}} = 999.9546 \approx c.$$

## 6.23 Parte 1 – Un modello ibrido

**Prezzi e quantità come variabili decisionali** Modelli diversi, come si vedrà, possono essere costruiti basandosi su quantità o prezzi. Qui si considera un esempio numerico istruttivo (Kuyumcu e Popescu, 2006) che mostra che prezzi e quantità non devono necessariamente essere variabili decisionali mutuamente esclusive.

**Domanda lineare per due beni complementari** Si considerino due articoli che seguono un modello di domanda lineare:

$$d_1 = 500 - p_1 - 5p_2, \quad d_2 = 10 - 0.01p_1 - 0.05p_2.$$

L'ispezione dei segni mostra che i due beni sono **complementi**, e non sostituti (ridurre  $p_2$  aumenta  $d_1$ ). Nel lavoro originale, i due beni sono camere standard e sale riunioni in un hotel, che chiaramente non sono sostituti. Questo aiuta a comprendere i valori numerici dei parametri.

**Vincoli di capacità** In pratica possono esistere limitazioni di capacità. In questo caso, la disponibilità è limitata a 250 e 6, rispettivamente. È quindi necessario modellare esplicitamente la domanda.

## 6.24 Ottimizzazione dei prezzi con domanda legata alle vendite

**Prima formulazione (non sensata)** Per trovare i prezzi ottimi si potrebbe considerare il seguente modello:

$$\max p_1 d_1 + p_2 d_2$$

soggetto a

$$\begin{aligned} d_1 &= 500 - p_1 - 5p_2, \\ d_2 &= 10 - 0.01p_1 - 0.05p_2, \\ p_1, p_2, d_1, d_2 &\geq 0, \\ d_1 &\leq 250, \quad d_2 \leq 6. \end{aligned}$$

Tuttavia, questo non è un modello particolarmente sensato.

**Soluzione della prima formulazione** La soluzione, in termini di prezzi ottimi, è:

$$p_1^* = 400, \quad p_2^* = 0.$$

Questi prezzi implicano le domande:

$$d_1^* = 100, \quad d_2^* = 6.$$

Si osserva che le sale riunioni sono offerte gratuitamente, cosa che può essere sensata per stimolare le vendite del bene complementare.

**Prezzi nulli o negativi per beni complementari** In alcuni casi estremi, il prezzo di un bene potrebbe anche essere negativo per incrementare le vendite del bene complementare. Questo può sembrare strano ma, in realtà, alcuni beni possono essere venduti sotto costo per ottenere ricavi dai beni complementari. Ad esempio, una stampante laser permette di vendere cartucce toner, e un rasoio permette di vendere lame.

**Effetti su capacità e ricavi** Il prezzo per le camere standard è piuttosto elevato e alcune camere restano inutilizzate, mentre la capacità è vincolante per le sale riunioni. Il ricavo complessivo è 40,000.

## 6.25 Che cosa non funziona nel primo modello?

**Collegamento rigido tra domanda e vendite** Il prezzo delle camere standard è elevato ed è tale che la domanda di sale riunioni è esattamente pari alla disponibilità.

**Inventory rationing** Una pratica comune nel revenue management è l'**inventory rationing**, ossia la restrizione delle vendite. Questo è lo strumento principale nel revenue management basato sulle quantità, come si vedrà.

## 6.26 Formulazione alternativa: disaccoppiare vendite e domanda

**Seconda formulazione (con razionamento)** Si può disaccoppiare la vendita dalla domanda risolvendo la seguente formulazione alternativa:

$$\max p_1 q_1 + p_2 q_2$$

soggetto a

$$\begin{aligned} q_1 &\leq 500 - p_1 - 5p_2, \\ q_2 &\leq 10 - 0.01p_1 - 0.05p_2, \\ p_1, p_2, q_1, q_2 &\geq 0, \\ q_1 &\leq 250, \quad q_2 \leq 6. \end{aligned}$$

Qui, la funzione di domanda lineare fornisce un limite superiore alla quantità offerta.

**Soluzione della seconda formulazione** La soluzione ottima è:

$$p_1^* = 250, \quad p_2^* = 0, \quad q_1^* = 250, \quad q_2^* = 6,$$

con ricavo pari a 62,500.

**Interpretazione della differenza tra le due formulazioni** Il problema della prima formulazione è che questa soluzione produrrebbe una domanda pari a 7.5 per il secondo bene, che non è realizzabile se la domanda è rigidamente legata alle vendite. Il disaccoppiamento tramite razionamento fornisce gradi di libertà aggiuntivi.

## 6.27 Parte 2 – Modelli di teoria dei giochi

**Problemi decisionali con più decisorî** Talvolta le decisioni di pricing non possono essere prese ignorando il contesto complessivo. In un contesto B2C, i consumatori strategici sono un problema. Le interazioni strategiche sono ancora più rilevanti in un contesto B2B, dove i produttori possono essere collegati da una supply chain in cui beni intermedi vengono scambiati e trasformati, oppure dove produttori e retailer che distribuiscono beni finali interagiscono. In tale contesto, i prezzi possono svolgere un ruolo importante come strumento per coordinare azioni e condividere rischi. Infine, che dire della competizione tra imprese? Occorre distinguere tra competizione in termini di prezzi o di quantità, e tra beni/servizi identici o differenziati. Questi problemi possono essere (almeno parzialmente) affrontati con modelli di teoria dei giochi.

**Ipotesi semplificative** Per semplicità, si considerano giochi stilizzati:

- ci sono solo due decisori (giocatori); ciascun giocatore ha un obiettivo (payoff) che vuole massimizzare e non esiste cooperazione;
- ciascun giocatore prende una sola decisione; quindi non si considerano giochi sequenziali con decisioni multiple nel tempo;
- si assume informazione completa e common knowledge: non c'è incertezza sui dati del problema né sui meccanismi che mappano decisioni in payoff; i due giocatori condividono la stessa visione del mondo e le regole del gioco, conoscono gli incentivi della controparte e ciascuno sa che l'altro possiede tutte le informazioni rilevanti.

**Problema centralizzato e interazione tra decisioni** Per avvicinarsi a una formalizzazione, si consideri il problema:

$$\max \pi_1(x_1; x_2) + \pi_2(x_1; x_2) \quad (4)$$

soggetto a

$$x_1 \in S_1, \quad x_2 \in S_2.$$

La funzione obiettivo (4) può essere interpretata come un profitto dipendente da due variabili decisionali,  $x_1$  e  $x_2$ , che devono appartenere ai rispettivi insiemi ammissibili  $S_1$  e  $S_2$ .

**Perché non si può decomporre** Anche se i vincoli su  $x_1$  e  $x_2$  sono separabili, non è possibile decomporre il problema complessivo, poiché le due decisioni interagiscono tramite le funzioni di profitto  $\pi_1(x_1; x_2)$  e  $\pi_2(x_1; x_2)$ . Risolvendo (4) si ottengono decisioni ottime  $x_1^*$  e  $x_2^*$  e il profitto totale ottimo:

$$\pi_{1+2}^* = \pi_1(x_1^*; x_2^*) + \pi_2(x_1^*; x_2^*).$$

Così facendo si assume un decisore unico che prende entrambe le decisioni, oppure una coppia di decisori cooperativi che scelgono  $x_1$  e  $x_2$  condividendo l'obiettivo di massimizzare la somma dei profitti.

**Caso non cooperativo** Nel caso realistico di due decisori non cooperativi con profitti  $\pi_1(x_1; x_2)$  e  $\pi_2(x_1; x_2)$ , il decisore 1 vorrebbe risolvere:

$$\max \pi_1(x_1; x_2) \quad (5)$$

soggetto a

$$x_1 \in S_1,$$

mentre il decisore 2 vorrebbe risolvere:

$$\max \pi_2(x_1; x_2) \quad (6)$$

soggetto a

$$x_2 \in S_2.$$

Questi due problemi, scritti così, non hanno senso: quale valore di  $x_2$  usare nel problema (5)? Quale valore di  $x_1$  usare nel problema (6)? È necessario chiarire come i due decisori muovono.

### Sequenziale vs. simultaneo

1. Una possibilità è che i decisori agiscano in modo sequenziale. Ad esempio, il decisore 1 sceglie  $x_1 \in S_1$  prima che il decisore 2 scelga  $x_2 \in S_2$ . In tal caso, il decisore 1 è il leader e il decisore 2 è il follower. Nel prendere la decisione, il decisore 1 può cercare di anticipare la reazione del decisore 2 a ciascun valore possibile di  $x_1$ .
2. Un'altra possibilità è che le due decisioni siano prese simultaneamente. In tal caso, servono strumenti concettuali per capire quali decisioni aspettarsi.

**Equilibrio e perdita di performance** La teoria dei giochi mira a fornire una previsione sensata di una soluzione di equilibrio  $(x_1^e; x_2^e)$ , che dipende dalle ipotesi sulla struttura del gioco. Qualunque equilibrio si ottenga, esso non può produrre un profitto totale maggiore di  $\pi_{1+2}^*$ , poiché necessariamente vale:

$$\pi_{1+2}^e = \pi_1(x_1^e; x_2^e) + \pi_2(x_1^e; x_2^e) \leq \pi_1(x_1^*; x_2^*) + \pi_2(x_1^*; x_2^*) = \pi_{1+2}^*.$$

Se questa diseguaglianza fosse violata,  $(x_1^*; x_2^*)$  non sarebbe soluzione ottima di (4). Questo significa che, decentralizzando le decisioni, il sistema complessivo è verosimilmente incapace di raggiungere la performance ottima globale.

## 6.28 Giochi con decisioni continue

**Equilibrio di Nash con azioni continue** Nel seguito si considera l'equilibrio di Nash nel caso in cui ogni giocatore disponga di un continuo di azioni possibili. Si analizza prima il comportamento di due imprese che competono in termini di quantità. Entrambe massimizzano il profitto, ma influenzano reciprocamente il prezzo di mercato tramite la quantità totale prodotta. Il prezzo è comune a entrambe, poiché si assume un prodotto perfettamente identico. Questa competizione si chiama competizione di Cournot; il caso in cui le imprese competono sui prezzi si chiama competizione di Bertrand. Nel caso di competizione sui prezzi, si distingue tra prodotti omogenei o differenziati.

## 6.29 Competizione di Cournot

**Impostazione del modello** Un gioco a mosse simultanee in cui le azioni sono quantità conduce all'equilibrio di Cournot–Nash. Ciò può essere rilevante in mercati, come l'energia, dove le imprese scelgono quantità e i prezzi sono determinati tramite meccanismi d'asta. Si considera un modello molto semplice in cui ciascuna impresa ha solo costo variabile:

$$TC_i(q_i) = c_i q_i, \quad i = 1, 2,$$

dove  $TC_i$  è il costo totale dell'impresa  $i$ ,  $c_i$  è il costo variabile, e  $q_i$  è la quantità prodotta dall'impresa  $i$ . La quantità totale sul mercato è  $Q = q_1 + q_2$  e influenza il prezzo secondo una funzione di domanda inversa lineare:

$$P(Q) = a - bQ, \quad a, b > 0, \quad a \geq c_i.$$

Questo modello implica che tutta la produzione venga venduta sul mercato. Il profitto dell'impresa  $i$  è:

$$\pi_i(q_1; q_2) = P(q_1 + q_2)q_i - TC_i(q_i) = [a - b(q_1 + q_2)]q_i - c_i q_i, \quad i = 1, 2.$$

**Funzioni di reazione (best response)** Assumendo decisioni simultanee, si può determinare l'equilibrio tramite le funzioni di best response  $R_i(q_j)$ . La condizione di stazionarietà per l'impresa 1 è:

$$\frac{\partial \pi_1(q_1; q_2)}{\partial q_1} = a - 2bq_1 - bq_2 - c_1 = 0 \Rightarrow R_1(q_2) = \frac{a - c_1}{2b} - \frac{1}{2}q_2. \quad (7)$$

Analogamente, per l'impresa 2:

$$R_2(q_1) = \frac{a - c_2}{2b} - \frac{1}{2}q_1. \quad (8)$$

**Sistema per l'equilibrio di Cournot** Per trovare l'equilibrio si individua l'intersezione delle due funzioni di risposta, cioè si risolve:

$$\begin{cases} q_1^c = R_1(q_2^c) \\ q_2^c = R_2(q_1^c) \end{cases}$$

(dove il superscritto  $c$  denota l'equilibrio di Cournot; le funzioni di risposta sono rette a pendenza negativa). Si risolve quindi il sistema lineare:

$$\begin{cases} q_1^c = \frac{a - c_1}{2b} - \frac{1}{2}q_2^c \\ q_2^c = \frac{a - c_2}{2b} - \frac{1}{2}q_1^c \end{cases}$$

ottenendo:

$$q_1^c = \frac{a - 2c_1 + c_2}{3b}, \quad q_2^c = \frac{a - 2c_2 + c_1}{3b}. \quad (9)$$

**Prezzo di equilibrio e profitti** Il prezzo di equilibrio risulta:

$$p^c = \frac{a + c_1 + c_2}{3}. \quad (10)$$

Il profitto di ciascuna impresa è:

$$\pi_i^c = (p^c - c_i)q_i^c = \left( \frac{a + c_i + c_j}{3} - c_i \right) \left( \frac{a - 2c_i + c_j}{3b} \right) = \frac{(a - 2c_i + c_j)^2}{9b} = b(q_i^c)^2. \quad (11)$$

Se un'impresa riduce il proprio costo, aumenta sia la quantità prodotta sia il profitto. Se le imprese hanno la stessa tecnologia ( $c_1 = c_2$ ), si ottiene una soluzione simmetrica  $q_1^c = q_2^c$ , come atteso.

## 6.30 Competizione di Bertrand

**Impostazione del problema** Si considerino due imprese che competono sui prezzi per un bene omogeneo, sotto l'ipotesi semplice di domanda e costi lineari. È necessario specificare la reazione dei consumatori ai prezzi. Con prodotto omogeneo, è sensato assumere che i consumatori acquistino dal venditore più economico; se i prezzi sono uguali, si assume che il mercato sia diviso equamente tra le due imprese. Non si considerano vincoli di capacità.

**Quantità venduta in funzione dei prezzi** La quantità venduta dall'impresa  $i$  (con  $i = 1, 2$  e  $j = 2, 1$  per indicare il concorrente) è:

$$q_i = \begin{cases} 0 & \text{se } p_i \geq \alpha/\beta, \\ 0 & \text{se } p_i > p_j, \\ \frac{\alpha - \beta p}{2} & \text{se } p_i = p_j \equiv p < \alpha/\beta, \\ \alpha - \beta p_i & \text{se } p_i < \min\{p_j, \alpha/\beta\}. \end{cases} \quad (12)$$

L'idea è che non ci sia domanda se il prezzo supera il prezzo limite  $\alpha/\beta$  oppure se è strettamente maggiore di quello del concorrente. Se i prezzi coincidono (e sono sotto il limite) la domanda viene divisa a metà; altrimenti l'impresa più economica cattura tutta la domanda.

**Definizione di equilibrio di Bertrand–Nash** Un equilibrio di Bertrand–Nash è una quadrupla di prezzi e quantità,

$$(p_1^b, p_2^b, q_1^b, q_2^b),$$

tale che:

1.  $p_1^b$  risolve  $\max_{p_1} \pi_1(p_1; p_2^b) = (p_1 - c_1)q_1$ , per  $p_2 = p_2^b$ ;
2.  $p_2^b$  risolve  $\max_{p_2} \pi_2(p_1^b; p_2) = (p_2 - c_2)q_2$ , per  $p_1 = p_1^b$ ;
3. le quantità  $q_1$  e  $q_2$  sono date dall'equazione (12).

Chiaramente, i prezzi non saranno fissati al di sotto dei costi marginali  $c_1$  e  $c_2$ , rispettivamente.

**Undercutting e (assenza di) equilibrio con costi diversi** Il concetto chiave è l’undercutting. Se  $c_1 < c_2$ , l’impresa 1 può undercut l’impresa 2 fissando un prezzo strettamente minore di  $p_2$ . In senso stretto non esiste equilibrio, poiché non c’è un massimo (ma solo un supremo). In effetti, c’è una discontinuità (a differenza di Cournot): una diminuzione infinitesima del prezzo può generare un salto della domanda per un’impresa.

**Equilibrio con costi marginali uguali** Si può mostrare che esiste equilibrio quando il costo marginale è lo stesso  $c$  per entrambe le imprese:

$$p_1 = p_2 = c, \quad q_1 = q_2 = \frac{\alpha - \beta c}{2}.$$

Il risultato si dimostra facilmente per assurdo: per qualunque altra configurazione dei prezzi, un’impresa avrebbe incentivo a deviare.

**Conseguenze e discreto passo di prezzo** Il risultato netto è che può non esserci profitto per nessuna impresa. Se i costi marginali sono diversi, per trovare un massimo (e un equilibrio) è necessario che i prezzi possano variare solo di un ammontare minimo  $\varepsilon$  (ad esempio un centesimo di euro). Se  $c_2 - c_1 > \varepsilon$ , allora l’impresa 1 può undercut impostando  $p_1 = p_2 - \varepsilon$ , e l’equilibrio risultante è:

$$p_2 = c_2, \quad p_1 = c_2 - \varepsilon, \quad q_1 = \alpha - \beta(c_2 - \varepsilon), \quad q_2 = 0.$$

**Osservazioni finali** Lasciando da parte le tecnicità, questa previsione è in parte in contrasto con i risultati empirici. Nella pratica, la struttura dei costi non è semplicemente lineare, esistono vincoli di capacità e i beni non sono realmente omogenei. Possono emergere ulteriori complicazioni, per cui non è detto che esista una sola impresa sul mercato. Guardando il risultato in modo opposto, esso suggerisce che le imprese dovrebbero differenziare le proprie offerte e sfruttare l’eterogeneità dei consumatori.

## 6.31 Parte 2 – Competizione di prezzo con beni non omogenei: modello di Hotelling

**Beni differenziati e scelta del consumatore** Per trattare beni non omogenei (differenziati), è necessario un modello di scelta del consumatore. Si consideri un modello unidimensionale delle preferenze dei consumatori. I beni prodotti da due imprese,  $A$  e  $B$ , sono caratterizzati da una feature, misurata su una scala da 0 a  $L$ . Per semplicità, si trascurano i costi di produzione. L’impresa  $A$  produce un bene con feature al livello  $a$ , e l’impresa  $B$  produce un bene con feature al livello  $L - b$ , a distanza  $b$  dal livello massimo. Questa rappresentazione geometrica è nota come *Hotelling’s linear street model*. Ogni consumatore ha un livello preferito per la feature, rappresentato dalla sua posizione sulla “strada”. Si assume che i consumatori siano uniformemente distribuiti sulla strada, dal livello 0 al livello  $L$ . Ogni consumatore ha una marca preferita, quella più vicina ai propri gusti. Se i due prezzi  $p_A$  e  $p_B$  sono uguali, ciascun consumatore acquista semplicemente la marca più vicina. Tuttavia, esiste un trade-off tra prezzo e soddisfazione.

**Costo di trasporto e utilità** Si immagini di esprimere la (dis)utilità di ciascun consumatore misurando un costo di “trasporto”  $\tau$ . Per un consumatore in posizione  $x$  sulla scala, l’utilità è:

$$u(x) = \begin{cases} -p_A - \tau|x - a| & \text{se sceglie la marca } A, \\ -p_B - \tau|x - (L - b)| & \text{se sceglie la marca } B. \end{cases}$$

Si cerca un consumatore critico, in posizione  $x_i$ , tale che  $a < x_i < L - b$ , indifferente tra le due marche:

$$-p_A - \tau(x_i - a) = -p_B - \tau(L - b - x_i),$$

da cui segue:

$$x_i = \frac{p_B - p_A}{2\tau} + \frac{L - b + a}{2}. \tag{13}$$

Come controllo di coerenza, quando i prezzi sono uguali,  $x_i$  è il punto medio tra le due posizioni delle marche. L'impresa  $A$  cattura tutta la domanda nell'intervallo  $[0, x_i]$ , quindi la sua domanda è  $x_i$ , mentre la domanda per la marca  $B$  è:

$$L - x_i = \frac{p_A - p_B}{2\tau} + \frac{L + b - a}{2}.$$

Ci si può chiedere se esista un equilibrio di Bertrand–Nash nei prezzi, a posizioni fissate.

**Problemi di massimizzazione dei profitti** L'impresa  $A$ , dato  $p_B$ , risolve:

$$\max_{p_A} \pi_A = \left( \frac{p_B p_A - p_A^2}{2\tau} + \frac{L - b + a}{2} \right) p_A,$$

mentre l'impresa  $B$ , dato  $p_A$ , risolve:

$$\max_{p_B} \pi_B = \left( \frac{p_A p_B - p_B^2}{2\tau} + \frac{L + b - a}{2} \right) p_B.$$

**Condizioni del primo ordine ed equilibrio nei prezzi** Le condizioni di stazionarietà del primo ordine sono:

$$\frac{\partial \pi_A}{\partial p_A} = \frac{p_B - 2p_A}{2\tau} + \frac{L - b + a}{2} = 0, \quad \frac{\partial \pi_B}{\partial p_B} = \frac{p_A - 2p_B}{2\tau} + \frac{L + b - a}{2} = 0,$$

che forniscono i prezzi di equilibrio:

$$p_A^e = \frac{\tau(3L - b + a)}{3}, \quad p_B^e = \frac{\tau(3L + b - a)}{3}.$$

**Profitto dell'impresa A** Il profitto per l'impresa  $A$  è:

$$\pi_A^e = x_i^e p_A^e = \frac{\tau(3L - b + a)^2}{18}.$$

**Caso simmetrico** Nel caso simmetrico  $a = b$  (che non implica che le posizioni coincidano), i prezzi sono uguali e il profitto si riduce a:

$$\frac{\tau L^2}{2}$$

per entrambe le imprese. Si osserva che, maggiore è il costo di trasporto, maggiori sono i prezzi. Ciò vale anche in contesti con switching costs.

**Esistenza di prezzi di equilibrio positivi** Non si deve dare per scontato che esista una coppia di prezzi di equilibrio strettamente positivi. Si può mostrare che:

1. se le marche sono omogenee (localizzate nello stesso punto), allora  $p_A^e = p_B^e = 0$  a causa dell'undercutting;
2. se le marche non sono troppo vicine, allora valgono gli equilibri precedenti; altrimenti, non esiste equilibrio (si veda la discussione successiva);
3. non esiste equilibrio nel gioco in cui sia le posizioni sia i prezzi sono variabili decisionali.

## 6.32 Esistenza di equilibrio nel modello di Hotelling

**Caso simmetrico e condizione su  $a$**  Si consideri il caso simmetrico  $a = b$ . Dalla discussione precedente si dovrebbe avere:

$$p_A^e = p_B^e = \tau L, \quad \pi_A^e = \pi_B^e = \frac{\tau L^2}{2}.$$

In realtà, si può dimostrare che ciò vale solo per  $a \leq L/4$ .

**Prezzo limite per catturare tutto il mercato** Si assuma  $p_B = \tau L$ , il prezzo di equilibrio, e si cambi  $p_A$ . L'impresa  $A$  cattura tutto il mercato se anche il consumatore in  $L - b$  sceglie  $A$ . Ciò accade se il prezzo  $p_A$  più il costo di trasporto è minore di  $p_B = \tau L$ . Osservando che la distanza tra  $A$  e  $B$  è  $L - 2a$ ,  $A$  cattura il mercato se:

$$p_A < \tau L - \tau(L - 2a) = 2a\tau = p_{A,\text{lim}}.$$

In questo caso, il profitto per  $A$  è una funzione lineare crescente di  $p_A$  e il suo valore limite è:

$$\pi_A^\Delta = p_{A,\text{lim}}L = 2a\tau L.$$

Per il prezzo limite c'è un'ambiguità, poiché i consumatori a destra di  $B$  sono indifferenti; si può assumere che quel segmento di mercato sia diviso in parti uguali. In ogni caso, esiste una discontinuità nel profitto di  $A$  in funzione del prezzo.

**Profitto quando il mercato è diviso** Per  $p_A > p_{A,\text{lim}}$ , il mercato si divide e, dalla (13), si ottiene:

$$\pi_A = \left[ \frac{p_B - p_A}{2\tau} + \frac{L - b + a}{2} \right] p_A = \left[ \frac{\tau L - p_A}{2\tau} + \frac{L}{2} \right] p_A = p_A L - \frac{p_A^2}{2\tau}.$$

All'equilibrio, quando  $p_A = \tau L$ , si ottiene:

$$\pi_A^e = \frac{\tau L^2}{2}.$$

Poi, quando  $p_A$  diventa abbastanza grande, l'impresa  $B$  cattura tutto il mercato e  $\pi_A$  scende a 0.

**Condizione di esistenza dell'equilibrio** Esiste un equilibrio (cioè esiste un massimo e non solo un supremo) se:

$$\pi_A^e \geq \pi_A^\Delta,$$

ossia:

$$\frac{\tau L^2}{2} \geq 2a\tau L \Rightarrow a \leq \frac{L}{4}.$$

Ciò significa che le imprese devono essere sufficientemente distanti per garantire l'esistenza di un equilibrio.

### 6.33 Giochi simultanei vs. giochi sequenziali

**Motivazione dei giochi sequenziali** Finora si è assunto che le due imprese competano simultaneamente e debbano prendere decisioni simili. Talvolta è più naturale assumere che uno dei due giocatori muova per primo. Ad esempio, un produttore può dover decidere il prezzo all'ingrosso, e un retailer la quantità acquistata (assumendo prezzo di vendita fissato).

**Da Cournot a von Stackelberg** Si consideri quindi che cosa accade nel gioco sulle quantità se si assume che l'impresa 1 (leader) scelga la quantità  $q_1$  prima dell'impresa 2 (follower). Diversamente dal gioco simultaneo, l'impresa 2 conosce la decisione di 1 prima di decidere; quindi ha informazione perfetta. L'analisi del gioco sequenziale porta all'equilibrio di von Stackelberg.

**Problema del leader** L'impresa 1 prende la decisione conoscendo la best response dell'impresa 2, data dall'equazione (8). Il problema del leader è:

$$\max_{q_1} \pi_1^s = P(q_1 + R_2(q_1))q_1 - c_1 q_1 = \left[ a - b \left( q_1 + \frac{a - c_2}{2b} - \frac{q_1}{2} \right) \right] q_1 - c_1 q_1,$$

dove il superscritto  $s$  si riferisce alla competizione di von Stackelberg.

**Quantità ottima del leader** Applicando la condizione di stazionarietà si ottiene:

$$q_1^s = \frac{a - 2c_1 + c_2}{2b} = \frac{3}{2} q_1^c.$$

L'impresa 1 produce di più nel gioco sequenziale rispetto a Cournot.

**Quantità del follower** Sostituendo in  $R_2(q_1)$  si ottiene:

$$q_2^s = \frac{a - c_2}{2b} - \frac{a - 2c_2 + c_1}{4b} = \frac{a - 3c_2 + 2c_1}{4b} = \frac{3}{4} q_2^c + \frac{c_1 - c_2}{4b}. \quad (15)$$

La produzione dell'impresa 2 è una frazione di quella del gioco di Cournot, più un termine positivo se l'impresa 1 è meno efficiente dell'impresa 2.

**Confronto dei profitti** È interessante confrontare i profitti delle due imprese in questo gioco. Ciò è semplice quando i costi marginali sono uguali; l'idea è illustrata con un esempio numerico.

### 6.34 Esempio

**Dati** Due imprese hanno lo stesso costo marginale,  $c_1 = c_2 = 5$ , e il mercato è caratterizzato dalla funzione prezzo/quantità:

$$P(Q) = 120 - Q.$$

Si confrontano tre casi:

1. le due imprese colludono e operano come un cartello (o come due rami di un monopolista);
2. le imprese non cooperano e muovono simultaneamente (gioco di Cournot);
3. le imprese non cooperano e muovono sequenzialmente (gioco di von Stackelberg).

**Caso 1: monopolio/cartello** Nel primo caso si lavora con l'output aggregato  $Q$ . Il monopolista risolve:

$$\max \pi^m = (120 - Q)Q - 5Q.$$

Applicando la condizione di stazionarietà:

$$120 - 2Q - 5 = 0 \Rightarrow Q^m = 57.50.$$

Il prezzo e il profitto risultano:

$$p^m = 120 - 57.5 = 62.50, \quad \pi_{1+2}^m = (62.50 - 5) \cdot 57.50 = 3306.25.$$

**Caso 2: Cournot (simultaneo)** Nel secondo caso, la soluzione (9) è simmetrica:

$$q_1^c = q_2^c = \frac{120 - 10 + 5}{3} = 38.33.$$

La produzione totale e il prezzo sono:

$$Q^c = 2 \cdot 38.33 = 76.77, \quad p^c = 120 - 76.77 = 43.33.$$

Il profitto di ciascuna impresa è:

$$\pi_1^c = \pi_2^c = (q_1^c)^2 = 1469.19.$$

Il profitto complessivo è:

$$\pi_{1+2}^c = 2 \cdot 1469.19 = 2938.89 < 3306.25 = \pi_{1+2}^m.$$

Il monopolista restringe l'output per aumentare il prezzo, ottenendo un profitto totale maggiore rispetto alla competizione di Cournot. La collusione produce quindi un profitto maggiore della competizione.

**Caso 3: von Stackelberg (sequenziale)** Nel gioco sequenziale, usando (14) e (15), si ottiene:

$$q_1^s = \frac{120 - 10 + 5}{2} = 57.5, \quad q_2^s = \frac{120 - 10 + 5}{4} = 28.75.$$

Rispetto al gioco simultaneo, l'output dell'impresa 1 aumenta mentre quello dell'impresa 2 diminuisce. La produzione totale e il prezzo sono:

$$Q^s = 57.5 + 28.75 = 86.25, \quad p^s = 120 - 86.25 = 33.75.$$

Il prezzo è inferiore rispetto ai due casi precedenti e la distribuzione del profitto è asimmetrica:

$$\pi_1^s = (33.75 - 5) \cdot 57.5 = 1653.13,$$

$$\pi_2^s = (33.75 - 5) \cdot 28.75 = 826.56,$$

$$\pi_{1+2}^s = 1653.13 + 826.56 = 2479.69.$$

Il profitto totale del gioco sequenziale è inferiore a quello del gioco simultaneo; tuttavia, il leader ha un vantaggio netto e il suo profitto è maggiore nel gioco sequenziale.

### 6.35 Parte 2 – Conviene sempre muovere per primi?

**Vantaggio del leader e limiti** L'esempio giocattolo precedente mostra che il privilegio di muovere per primi può dare un vantaggio al leader. Data la struttura del gioco, è facile vedere che il leader nel gioco sequenziale non può fare peggio che nel gioco simultaneo; infatti, potrebbe comunque produrre la stessa quantità del gioco di Cournot. Tuttavia, ciò non vale in generale. In particolare, quando esistono asimmetrie informative o elementi casuali, la scelta del leader, o il suo esito in presenza di incertezza, può fornire al follower informazione utile. L'esempio seguente mostra che essere i primi a muovere non è sempre desiderabile.

**Esempio: battle of the sexes (Juliet muove per prima)** Si consideri il battle of the sexes, assumendo che Juliet abbia il privilegio di muovere per prima.

		Romeo: Horror	Romeo: Shopping
		(1, 3)	(0, 0)
Juliet: Horror	Juliet: Horror	(1, 3)	(0, 0)
	Juliet: Shopping	(0, 0)	(3, 1)

Qualunque sia la scelta di Juliet, Romeo sceglierà la mossa che gli consente di godere della sua compagnia. Di conseguenza, Juliet sceglierà certamente *shopping* ed è certamente felice di muovere per prima.

**Un caso diverso: Morticia** La situazione è molto diversa per i payoff seguenti:

		Romeo: Cinema	Romeo: Restaurant
		(5, -100)	(0, 1)
Morticia: Cinema	Morticia: Cinema	(5, -100)	(0, 1)
	Morticia: Restaurant	(0, 1)	(5, -100)

In questo caso, Romeo è indifferente tra andare al cinema o al ristorante. Ciò che teme davvero è una serata con Morticia. È facile vedere che questo gioco non ha un equilibrio di Nash, poiché uno dei due giocatori ha sempre incentivo a deviare. Un equilibrio può essere trovato ammettendo strategie miste, in cui i giocatori selezionano un'azione secondo una distribuzione di probabilità, legata all'incertezza sulla mossa del concorrente. Qui non si considerano strategie miste, ma il punto importante è che, in questo caso, nessun giocatore vorrebbe muovere per primo.

**Interpretazione economica dei due giochi** Si è notato che la prima versione del battle of the sexes è un gioco di coordinamento stilizzato per due imprese che devono adottare uno standard comune; nella seconda versione, un'impresa vuole adottare lo stesso standard del concorrente, mentre l'altra impresa vorrebbe selezionarne uno differente.

## 6.36 Pricing e double marginalization

**Impostazione B2B: produttore e retailer** Come applicazione interessante al pricing, si consideri un contesto B2B, in cui un produttore e un retailer interagiscono tramite prezzi (si veda Tirole, 2003, Capitolo 4). È interessante confrontare decisioni di prezzo e profitti complessivi della supply chain in due contesti:

1. impresa verticalmente integrata, che gestisce produzione e distribuzione (un solo decisore prende tutte le decisioni);
2. schema decentralizzato, in cui il produttore decide il prezzo all'ingrosso e il retailer decide il prezzo di mercato.

In entrambi i casi si considera un mercato con funzione di domanda lineare:

$$d(p) = 1 - p,$$

dove  $p$  è il prezzo di mercato, e si assume una struttura di costo lineare con costo marginale  $c < 1$ .

**Caso 1: impresa verticalmente integrata** Il problema complessivo dell'impresa integrata è:

$$\max_p (p - c)(1 - p).$$

Risolvendo, si ottengono prezzo ottimo, domanda e profitto:

$$p_{vi}^* = \frac{1+c}{2}, \quad d_{vi}^* = \frac{1-c}{2}, \quad \pi_{vi}^* = \frac{(1-c)^2}{4},$$

rispettivamente.

**Caso 2: supply chain decentralizzata** Se la catena non è integrata, ciascun attore imposta un prezzo. Si assuma che il produttore sia il leader in un gioco sequenziale e fissi un prezzo all'ingrosso  $p_w$ . Serve la funzione di risposta del retailer, che imposta il prezzo di mercato  $p_m$  risolvendo, dato  $p_w$ :

$$\max_{p_m} (p_m - p_w)(1 - p_m).$$

Questo produce la best response:

$$R_m(p_w) = \frac{1 + p_w}{2}.$$

**Problema del produttore e prezzo all'ingrosso di equilibrio** Il problema del produttore diventa:

$$\max_{p_w} (p_w - c) \left[ 1 - \frac{1 + p_w}{2} \right],$$

che fornisce il prezzo di equilibrio:

$$p_{w;dec}^* = \frac{1+c}{2}.$$

**Prezzo retail e domanda nel caso decentralizzato** Il prezzo retail è maggiore che nel caso integrato (si ricordi  $c < 1$ ):

$$p_{m;dec}^* = \frac{3+c}{4},$$

e la domanda di mercato è più piccola:

$$d_{dec}^* = \frac{1-c}{4}.$$

**Profitto complessivo nel caso decentralizzato** Il profitto complessivo del controllo decentralizzato è la somma dei due profitti:

$$\pi_{dec}^* = (p_{w;dec}^* - c) d_{dec}^* + (p_{m;dec}^* - p_{w;dec}^*) d_{dec}^* = \frac{(1-c)^2}{8} + \frac{(1-c)^2}{16} = \frac{3(1-c)^2}{16} < \pi_{vi}^*.$$

**Double marginalization e coordinamento della supply chain** Questo tipo di problema è noto come **double marginalization** ed è dovuto al fatto che entrambi i giocatori applicano un markup al costo marginale che osservano, e tale costo percepito cresce lungo la catena. In un contesto più realistico, si dovrebbe considerare anche l'incertezza della domanda, che introduce temi di risk sharing. In tal caso, politiche di prezzo (two-part tariffs e buyback contracts) possono essere utilizzate per superare il problema e coordinare la supply chain.

## 6.37 Parte 3 – Modelli di scelta discreta

### 6.38 Conjoint analysis

**Idea di base e utilizzi** La *conjoint analysis* è un metodo ampiamente utilizzato nel marketing (e in altri campi) per progettare prodotti e servizi, valutare l'impatto del prezzo e costruire simulatori di mercato. L'idea di base è che i consumatori reagiscono ad alternative rappresentate da combinazioni di valori degli attributi. Esistono alcune varianti della conjoint analysis, ma le principali sono:

- *ratings-based conjoint analysis* (quella tradizionale);
- *choice-based conjoint analysis* (la più comune).

**Modello a rating** Il rating di preferenza  $Y_i$  per l'alternativa  $i$  può essere modellato come:

$$Y_i = \sum_{k=1}^m U_k(x_{ik}) + \varepsilon_i,$$

dove ciascuna funzione  $U_k$  cattura la *partworth* per l'attributo  $k$ , ed  $\varepsilon_i$  è un termine di errore. Si noti che chiedere ai consumatori di dare priorità agli attributi è tipicamente inutile (“tutto è importante!”). Chiedere un rating  $Y_i$  può essere una scelta migliore.

**Limiti e interazioni** La conjoint analysis classica inferisce le partworth sulla base dei rating forniti dai consumatori per un insieme di alternative (stimoli, profili, ecc.). Il carattere additivo del modello è un chiaro limite, ma esistono modi per introdurre interazioni tra attributi, che possono essere rilevanti per valori specifici (ad esempio “rosso” e “Ferrari”).

**Dati di scelta: revealed vs. stated preferences** Tuttavia, assegnare rating non è il modo in cui i consumatori prendono decisioni. Per costruire un modello di preferenza migliore, si devono raccogliere dati da scelte effettive tra un insieme di alternative (inclusa l'opzione di non acquistare). Si possono avere:

- *revealed preference data*, dove si osservano scelte reali di mercato;
- *stated preference data*, dove si analizza la reazione di un panel di consumatori a un insieme di stimoli accuratamente progettati (alternative ipotetiche).

La *choice-based conjoint analysis* si basa su stated preferences, poiché offre grande flessibilità nella generazione degli stimoli e nella raccolta dati (al costo di una possibile perdita di realismo).

**Design degli esperimenti e stima dei modelli** Un ruolo chiave è svolto da un'attenta progettazione degli esperimenti (design ortogonali, ecc.), per massimizzare l'informazione raccolta senza imporre un carico cognitivo eccessivo all'utente (troppe alternative per scelta e troppe scelte). Oltre alle questioni classiche di design of experiments, è importante:

- scegliere (ed eventualmente combinare) attributi e livelli;
- considerare proibizioni (combinazioni prive di senso);
- evitare profili incoerenti (prestazioni migliori a prezzo più basso);
- gestire con attenzione variabili ordinali.

Un modello tipico stimabile tramite CBC è un modello *multinomial logit* (MNL). Esistono versioni sofisticate di CBC, basate su modelli Bayesiani gerarchici, per tenere conto dell'eterogeneità dei consumatori. Valutando le partworth utilities, si può prevedere la market share di beni non attualmente offerti, così come l'impatto delle decisioni di prezzo.

### 6.39 Microfondazioni dei modelli MNL: utilità casuali

**Utilità rappresentativa e componente casuale** Si assume che il consumatore scelga uno tra  $n$  beni, in base all'utilità di ciascuna alternativa. L'utilità del bene  $i \in [n]$  è  $V_i + \varepsilon_i$ , ossia somma di una componente deterministica e una componente casuale. La componente deterministica è spesso detta *representative utility*. L'alternativa  $i$  è scelta se:

$$V_i + \varepsilon_i > V_j + \varepsilon_j, \quad \forall j \neq i.$$

Quindi, la probabilità di scelta dell'alternativa  $i$  è:

$$\pi_i = \mathbb{P}\{V_i + \varepsilon_i > V_j + \varepsilon_j, \forall j \neq i\} = \mathbb{P}\{\varepsilon_i > \varepsilon_j + V_j - V_i, \forall j \neq i\}.$$

**Probabilità condizionata e integrazione** Assumendo indipendenza tra le componenti casuali, condizionatamente a  $\varepsilon_i$ , si ha:

$$\pi_i | \varepsilon_i = \prod_{j \neq i} F_j(\varepsilon_i + V_i - V_j),$$

dove  $F_j$  è la DF di  $\varepsilon_j$ . Quindi:

$$\pi_i = \int_{-\infty}^{+\infty} \prod_{j \neq i} F_j(\varepsilon_i + V_i - V_j) f_i(s) ds,$$

dove  $f_i$  è la PDF di  $\varepsilon_i$ .

**Caso i.i.d. Gumbel** Si assume ora che tutte le componenti casuali siano i.i.d. e seguano una distribuzione di Gumbel, con DF e PDF:

$$F(x) = \exp(-e^{-x}), \quad f(x) = e^{-x} \exp(-e^{-x}).$$

Allora:

$$\pi_i = \int_{-\infty}^{+\infty} \prod_{j \neq i} \exp(-e^{-(s+V_i-V_j)}) e^{-s} \exp(-e^{-s}) ds.$$

Per calcolare l'integrale, si osserva che  $s+V_i-V_j = s$  per  $j = i$ , quindi si può includere il fattore  $\exp(-e^{-s})$  nel prodotto e raccogliere i termini che non dipendono da  $j$ :

$$\pi_i = \int_{-\infty}^{+\infty} \prod_{j=1}^n \exp(-e^{-(s+V_i-V_j)}) e^{-s} ds = \int_{-\infty}^{+\infty} \exp\left(-\sum_{j=1}^n e^{-(s+V_i-V_j)}\right) e^{-s} ds$$

$$= \int_{-\infty}^{+\infty} \exp\left(-e^{-s} \sum_{j=1}^n e^{-(V_i - V_j)}\right) e^{-s} ds.$$

Si applica ora un cambio di variabile  $t = e^{-s}$ , così che  $-e^{-s}ds = dt$  e i limiti inferiore e superiore diventano rispettivamente  $+\infty$  e 0.

**Derivazione del modello MNL** Correggendo per il cambio di segno:

$$\pi_i = \int_0^1 \exp\left(-t \sum_{j=1}^n e^{-(V_i - V_j)}\right) dt = \frac{-\exp\left(-t \sum_j e^{-(V_i - V_j)}\right)}{\sum_j e^{-(V_i - V_j)}} \Big|_0^1 = \frac{1}{\sum_j e^{-(V_i - V_j)}} = \frac{e^{V_i}}{\sum_j e^{V_j}},$$

che corrisponde al modello MNL.

**Assunzione IIA** Il modello MNL è comune nel machine learning e nella classificazione ed è relativamente semplice da gestire in modelli di ottimizzazione (ad esempio decisioni di assortimento ottimo). Tuttavia, ha limitazioni definite, in particolare l'assunzione IIA (*Independence from Irrelevant Alternatives*). Si considerino gli odds relativi di scegliere gli item  $i$  e  $k$ :

$$\frac{\pi_i}{\pi_k} = \frac{e^{V_i}/\sum_j e^{V_j}}{e^{V_k}/\sum_j e^{V_j}} = e^{V_i - V_k},$$

che non dipendono da nessun'altra alternativa. Quindi gli odds relativi non cambiano quando nuove alternative vengono aggiunte al choice set.

**Esempio (red–blue bus problem)** Si consideri una decisione sul modo di viaggio tra auto e bus (un bus blu), assumendo utilità rappresentative uguali, quindi  $\pi_c = \pi_{bb} = 1/2$ . Si introduca ora un bus rosso, per cui è ragionevole assumere  $\pi_{rb} = \pi_{bb}$ . Con un modello MNL si ottiene  $\pi_c = \pi_{rb} = \pi_{bb} = 1/3$ , mentre sarebbe più ragionevole assumere una suddivisione della probabilità di scegliere l'autobus, quindi  $\pi_c = 1/2$  e  $\pi_{rb} = \pi_{bb} = 1/4$ . I modelli *nested logit* sono un possibile approccio per aggirare l'IIA, quando necessario. Se si assume che la componente casuale dell'utilità sia normalmente distribuita, si ottiene un modello *probit*.

## 6.40 Supplemento tecnico – EVT e massimi a blocchi

**Distribuzione dei massimi a blocchi** Nella *Extreme Value Theory* (EVT), tra le altre cose, si studia la distribuzione dei massimi a blocchi. Si consideri un campione i.i.d. di taglia  $n$  da una distribuzione con funzione di distribuzione  $F_X(x)$ . Qual è la distribuzione di  $M_n = \max\{X_1, \dots, X_n\}$ ? Che cosa accade quando  $n \rightarrow \infty$ ? Per indipendenza:

$$F_{\max}(x) = \mathbb{P}\{M_n \leq x\} = \mathbb{P}\{X_1 \leq x, \dots, X_n \leq x\} = \mathbb{P}\{X_1 \leq x\} \cdots \mathbb{P}\{X_n \leq x\} = F_X(x)^n.$$

**Esempio: uniforme** Per  $U \sim U(0, 1)$ :

$$F_U(x) = x \Rightarrow F_{\max}(x) = x^n, \quad x \in [0, 1].$$

Quando  $n \rightarrow \infty$ , la distribuzione collassa in una distribuzione degenere con massa di probabilità concentrata in  $x = 1$ . Inoltre, in questo caso:

$$\mathbb{E}[M_n] = \frac{n}{n+1}.$$

Un caso degenere come questo, dovuto al supporto limitato, non è particolarmente interessante. Se si considera una variabile casuale con supporto illimitato verso l'alto, la distribuzione si sposta verso  $+\infty$ . Quindi?

**Normalizzazione: confronto con il CLT** Nel caso della somma  $S_n = \sum_{k \in [n]} X_k$ , rilevante per il teorema del limite centrale, si normalizza definendo sequenze  $a_n$  e  $b_n$ :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \frac{S_n - a_n}{b_n} \leq x \right\} = \Phi(x), \quad a_n = n\mathbb{E}(X_1), \quad b_n = \sqrt{n \text{Var}(X_1)}.$$

**Normalizzazione per i massimi** Per ottenere una distribuzione limite per i massimi a blocchi, si introducono sequenze normalizzanti  $d_n$  e  $c_n$  e si verifica se:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \frac{M_n - d_n}{c_n} \leq x \right\} = \lim_{n \rightarrow \infty} F_X^n(c_n x + d_n) = H(x), \quad (16)$$

per una funzione di distribuzione non degenere  $H(x)$ .

**Esempio: esponenziale** Si consideri una distribuzione esponenziale con rate  $\beta > 0$ , per cui:

$$F(x) = 1 - e^{-\beta x}, \quad x \geq 0.$$

Scegliendo  $c_n = 1/\beta$  e  $d_n = (\log n)/\beta$  si ottiene:

$$F_X^n(c_n x + d_n) = \left( 1 - \frac{1}{n} e^{-x} \right)^n, \quad x \geq -\log n,$$

e:

$$\lim_{n \rightarrow \infty} F_X^n(c_n x + d_n) = \exp(-e^{-x}).$$

**Definizione: distribuzione GEV** La funzione di distribuzione della *generalized extreme value* (GEV) standard è:

$$H_\xi(x) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}), & \xi \neq 0, \\ \exp(-e^{-x}), & \xi = 0, \end{cases}$$

dove  $1 + \xi x > 0$ . Introducendo un parametro di location  $\mu \in \mathbb{R}$  e un parametro di scala  $\sigma > 0$ , si ottiene la famiglia:

$$H_{\xi;\mu;\sigma}(x) := H_\xi \left( \frac{x - \mu}{\sigma} \right).$$

Il parametro  $\xi$  definisce la forma di una famiglia di distribuzioni simili (dello stesso tipo). Si ricorda che quando due variabili casuali  $V$  e  $W$  differiscono in distribuzione per un parametro di scala  $a > 0$  e un parametro di location  $b \in \mathbb{R}$ , cioè  $V \stackrel{d}{=} aW + b$ , allora hanno lo stesso tipo (sono simili).

**Interpretazione del parametro di forma** La figura mostra densità per alcuni valori del parametro di forma (assumendo  $\mu = 0, \sigma = 1$ ):

- se  $\xi > 0$ , si ha una distribuzione di Fréchet;
- se  $\xi = 0$ , si ha una distribuzione di Gumbel;
- se  $\xi < 0$ , si ha una distribuzione di Weibull.

La distribuzione di Weibull ha un estremo destro finito. La distribuzione di Fréchet presenta una decrescita più lenta della Gumbel. Nota: la distribuzione di Gumbel, tra le altre cose, svolge un ruolo chiave nei modelli di scelta discreta.

**Dominio di attrazione e teorema FTG** Se la condizione (16) vale per una  $H$  non degenere, si dice che la funzione di distribuzione  $F$  appartiene al massimo dominio di attrazione di  $H$ :  $F \in MDA(H)$ . Nel caso della distribuzione esponenziale, si ha  $F \in MDA(H_0)$ .

**Teorema (Fisher–Tippett–Gnedenko)** Se  $F \in MDA(H)$  per una  $H$  non degenere, allora  $H$  deve essere una distribuzione GEV di tipo  $H_\xi$  per qualche valore di  $\xi$ .

## 7 CAP 7 – Il principio della programmazione dinamica

### 7.1 Che cos’è la programmazione dinamica

**Definizione di programmazione dinamica** La **programmazione dinamica (Dynamic Programming, DP)** non è un algoritmo come l’algoritmo del simplex per la programmazione lineare. È piuttosto un **principio notevolmente generale e flessibile** che può essere utilizzato per progettare una vasta gamma di algoritmi di ottimizzazione.

**Idea centrale** L’idea centrale si basa sulla **decomposizione di un problema decisionale dinamico multistadio** in una **sequenza di problemi più semplici a singolo stadio**.

**Problemi multistadio e multiperiodo** Un problema decisionale multistadio non deve essere confuso con un problema multiperiodo. Si consideri un problema a orizzonte finito in cui si devono selezionare  $T$  decisioni  $x_t$ , applicate in una sequenza di istanti temporali  $t = 0, 1, 2, \dots, T - 1$ . La domanda fondamentale è: *quando vengono prese queste decisioni?*

**Approccio open-loop e closed-loop** In un problema statico multiperiodo, tutte le decisioni vengono prese al tempo  $t = 0$ , come è naturale in un problema deterministico (**approccio open-loop**). In un problema stocastico, invece, si osserva nel tempo una traiettoria campionaria dei fattori di rischio rilevanti e si adattano le decisioni lungo il percorso.

**Strategia decisionale** Ciò che serve è una **strategia**, ovvero una **regola per prendere decisioni dopo aver osservato le realizzazioni casuali (approccio closed-loop)**.

### 7.2 Ambito di applicazione della programmazione dinamica

**Limiti della programmazione dinamica** La programmazione dinamica non è un approccio universale, poiché la sua applicazione richiede una struttura specifica nel modello del sistema, che deve essere **markoviana**.

**Classi di problemi trattabili** Nonostante ciò, la programmazione dinamica può essere applicata a una vasta gamma di problemi:

- problemi continui e discreti, dove la natura discreta o continua può riguardare:
  - variabili di stato;
  - variabili decisionali;
  - rappresentazione del tempo;
- problemi deterministici e stocastici;
- problemi a orizzonte finito e infinito.

**Dimensione del problema** All’interno di questa ampia collezione di problemi, può essere necessario risolvere problemi a dimensione finita o infinita. Per alcune strutture, la programmazione dinamica rappresenta **l’unico approccio risolutivo praticabile**.

## 7.3 Problemi decisionali dinamici

**Modelli a tempo discreto** Si considerano modelli a tempo discreto ed è necessario chiarire la distinzione tra **istanti temporali** e **intervalli di tempo**.

**Decisioni e intervalli** Le decisioni vengono prese agli istanti temporali, ma possono essere implementate lungo un intervallo di tempo, non necessariamente in modo istantaneo. Inoltre, le decisioni devono essere prese prima di osservare la realizzazione dei fattori di rischio, che si manifestano nell'intervallo temporale successivo.

**Convenzione temporale** Si adotta la seguente convenzione:

- gli istanti temporali sono indicizzati da  $t = 0, 1, 2, \dots$ ; in tali istanti si osserva lo stato del sistema e si prende una decisione;
- l'intervallo di tempo  $t$  è l'intervallo tra gli istanti  $t - 1$  e  $t$ ; dopo aver applicato la decisione presa all'istante  $t - 1$ , il sistema evolve durante l'intervallo successivo e raggiunge un nuovo stato all'istante  $t$ ;
- durante tale intervallo può realizzarsi un input esterno, eventualmente casuale, che influenza la transizione di stato.

## 7.4 Orizzonte temporale e dinamica del sistema

**Orizzonte finito e infinito** Un problema può avere un orizzonte finito, definito sugli istanti  $t = 0, 1, \dots, T$ , oppure un orizzonte infinito, definito sugli istanti  $t = 0, 1, \dots$ . Il primo istante temporale è  $t = 0$ , mentre il primo intervallo temporale è indicizzato da  $t = 1$ . In un problema a orizzonte finito, vi sono  $T$  intervalli di tempo e  $T + 1$  istanti temporali.

**Equazione di transizione di stato** La dinamica del sistema è rappresentata da un'equazione di transizione di stato del tipo:

$$s_{t+1} = g_{t+1}(s_t, x_t, \xi_{t+1}),$$

dove:

- $g_{t+1}$  è la funzione di transizione sul periodo  $t + 1$ ;
- $s_t$  è il vettore delle variabili di stato all'istante  $t$ , che riassumono tutta l'informazione rilevante del passato;
- $s_0$  è lo stato iniziale, tipicamente noto, mentre  $s_T$  è lo stato terminale;
- $x_t$  è il vettore delle variabili decisionali all'istante  $t$ , dette anche variabili di controllo;
- $\xi_{t+1}$  è un fattore esogeno che si realizza durante l'intervallo  $t + 1$ .

## 7.5 Variabili di stato e processi

**Tipologie di variabili di stato** Si distinguono diverse tipologie di variabili di stato:

- **variabili di stato fisiche**, che descrivono risorse fisiche o finanziarie e sono direttamente influenzate dalle decisioni;
- **variabili di stato informative**, come il prezzo di un'azione finanziaria, non influenzate dall'azione di un singolo decisore;
- **variabili di stato di credenza**, rilevanti in problemi con incertezza sui parametri del modello.

**Processi decisionali e stocastici** Si definiscono:

- il processo dei dati  $(\xi_t)_{t \geq 1}$ ;
- il processo decisionale  $(x_t)_{t \geq 0}$ ;
- il processo di stato  $(s_t)_{t \geq 0}$ ;

che possono essere deterministici o stocastici.

## 7.6 Informazione e non anticipatività

**Disponibilità dell'informazione** Nei problemi decisionali dinamici sotto incertezza, due aspetti sono cruciali:

1. come modellare la disponibilità dell'informazione;
2. come sono correlati i processi dei dati, dello stato e delle decisioni.

**Storia osservata** Si introduce la notazione:

$$x[t] := (x_0, x_1, \dots, x_t), \quad \xi[t] := (\xi_1, \xi_2, \dots, \xi_t).$$

**Politiche non anticipative** La decisione  $x_t$  può dipendere solo dalle realizzazioni dei fattori di rischio fino all'istante  $t$  incluso. Una politica decisionale implementabile è detta **non anticipativa**.

## 7.7 Dipendenza temporale dei fattori di rischio

**Struttura del processo dei dati** I vettori casuali del processo dei dati possono presentare diversi livelli di dipendenza:

- indipendenza interstadio;
- dipendenza markoviana;
- dipendenza dall'intera storia osservata.

**Endogeneità ed esogeneità** È inoltre necessario specificare se la distribuzione dei fattori di rischio dipende dal tempo o dallo stato e dalle decisioni.

**Obiettivo additivo** L'obiettivo è ottimizzare una funzione obiettivo additiva nel tempo, pari alla somma dei costi o dei ricavi generati in ciascun istante.

## 7.8 Problemi a orizzonte finito scontati

**Formulazione generale** Un problema stocastico a orizzonte finito  $T$  può essere formulato come:

$$\text{opt } \mathbb{E}_0 \left[ \sum_{t=0}^{T-1} \gamma^t f_t(s_t, x_t) + \gamma^T F_T(s_T) \right],$$

dove opt indica una minimizzazione o massimizzazione.

**Interpretazione dei termini** La funzione obiettivo include:

- il contributo immediato  $f_t(s_t, x_t)$ ;
- il contributo terminale  $F_T(s_T)$ .

**Fattore di sconto** Il fattore di sconto  $\gamma \in (0, 1)$  è tipico nelle applicazioni finanziarie, anche se non è strettamente necessario in un problema a orizzonte finito.

## 7.9 Contributi stocastici e apprendimento

**Contributo dipendente dal rischio** Il contributo immediato può dipendere dalla realizzazione del fattore di rischio:

$$h_t(s_t, x_t, \xi_{t+1}).$$

**Aspettazione condizionata** In tal caso, si può definire:

$$f_t(s_t, x_t) = \mathbb{E}_t [h_t(s_t, x_t, \xi_{t+1})].$$

**Difficoltà pratiche** In pratica, il calcolo dell'aspettazione può essere difficile o impossibile. In tali casi, si dispone solo di campioni osservabili tramite simulazione Monte Carlo o sperimentazione online, come avviene nel **reinforcement learning**.

## 7.10 Problemi a orizzonte infinito

**Problemi scontati a orizzonte infinito** Un problema di programmazione dinamica scontato a orizzonte infinito ha la forma:

$$\text{opt } \mathbb{E}_0 \left[ \sum_{t=0}^{\infty} \gamma^t f(s_t, x_t) \right], \quad (6)$$

dove la sommatoria è ben definita se tutti i contributi sono positivi e limitati e si utilizza un fattore di sconto appropriato  $\gamma < 1$ .

**Ruolo del fattore di sconto** La programmazione dinamica scontata è naturale nelle applicazioni di business ed economia. Talvolta il fattore di sconto non ha una reale giustificazione economica ed è utilizzato solo come espediente per rendere trattabile un problema a orizzonte infinito.

**Criterio del costo medio** In alcune applicazioni ingegneristiche si preferisce considerare una media nel tempo:

$$\text{opt } \lim_{T \rightarrow \infty} \mathbb{E}_0 \left[ \frac{1}{T} \sum_{t=0}^{T-1} f(s_t, x_t) \right]. \quad (7)$$

Questo conduce a formulazioni di programmazione dinamica basate sul contributo medio per stadio.

**Orizzonte non definito** Esistono anche problemi con orizzonte non definito, nei quali il processo decisionale termina quando viene raggiunto uno stato obiettivo.

## 7.11 Politiche decisionali

**Vincolo di fattibilità delle decisioni** Ad ogni istante temporale occorre selezionare una decisione  $x_t$ . Si può introdurre un vincolo astratto del tipo:

$$x_t \in X(s_t), \quad (8)$$

che impone che la decisione al tempo  $t$  appartenga a un insieme ammissibile  $X$ , eventualmente dipendente dallo stato corrente  $s_t$ .

**Caso deterministico e stocastico** In un contesto deterministico, questo vincolo sarebbe sufficiente, poiché al tempo  $t = 0$  si dovrebbe individuare una sequenza di decisioni  $(x_t)_{t \geq 0}$  da implementare nel tempo. Nel caso stocastico, tuttavia, entra in gioco un vincolo fondamentale.

**Non anticipatività** Le decisioni devono essere **non anticipative**, cioè non possono dipendere da informazioni future non ancora osservate.

**Politiche in feedback** Una politica decisionale chiusa e non anticipativa è naturalmente espressa nella forma:

$$x_t = \mu_t(s_t) \in X(s_t), \quad (9)$$

dove  $\mu_t(\cdot)$  è una funzione che associa allo stato al tempo  $t$  una decisione ammissibile.

**Politica come sequenza di funzioni** Una politica può essere vista come una sequenza di funzioni:

$$\mu \equiv (\mu_0(\cdot), \mu_1(\cdot), \dots, \mu_{T-1}(\cdot)), \quad (10)$$

una per ciascun istante temporale in cui è richiesta una decisione.

**Politiche stazionarie** Nel caso di problemi a orizzonte infinito, si cercano politiche **stazionarie**, caratterizzate dalla stessa funzione  $\mu(\cdot)$  in ogni stadio.

**Formulazione del problema in termini di politiche** Una formulazione più precisa del problema a orizzonte finito è:

$$\text{opt}_{\mu \in \mathcal{M}} \mathbb{E}_0 \left[ \sum_{t=0}^{T-1} \gamma^t f_t(s_t, \mu_t(s_t)) + \gamma^T F_T(s_T) \right], \quad (11)$$

dove  $\mathcal{M}$  è l'insieme delle politiche ammissibili, tali che  $x_t = \mu_t(s_t) \in X(s_t)$  per ogni istante temporale.

**Politiche deterministiche e randomizzate** Sono state introdotte politiche deterministiche. Esistono tuttavia casi in cui politiche randomizzate possono essere necessarie o utili.

**Vincoli probabilistici sugli stati** Le politiche randomizzate possono essere richieste per gestire vincoli probabilistici sugli stati:

$$\mathbb{P}\{s_t \in G\} \geq 1 - \alpha,$$

dove  $\alpha$  è una probabilità accettabile di violazione del vincolo.

**Esplosione e sfruttamento** Le politiche randomizzate risultano inoltre necessarie nel contesto del trade-off tra esplorazione e sfruttamento nel reinforcement learning.

## 7.12 Un esempio: il lot-sizing dinamico a singolo prodotto

**Formulazione del problema** Si consideri il problema di lot-sizing non capacitato a singolo prodotto:

$$\min \sum_{t=0}^{T-1} [cx_t + \varphi \cdot \delta(x_t)] + \sum_{t=1}^T hI_t \quad (12)$$

soggetto a:

$$\begin{aligned} I_{t+1} &= I_t + x_t - d_{t+1}, & t = 0, 1, \dots, T-1, \\ x_t, I_t &\geq 0. \end{aligned} \quad (13)$$

**Significato delle variabili** Dove:

- $x_t$  è la quantità ordinata e consegnata immediatamente agli istanti  $t = 0, 1, \dots, T-1$ ;
- $I_t$  è il livello di inventario disponibile agli istanti  $t = 0, 1, \dots, T$ ;
- $d_t$  è la domanda durante gli intervalli temporali  $t = 1, \dots, T$ .

**Struttura dei costi** L'obiettivo è soddisfare la domanda al costo minimo, che comprende:

- un costo di mantenimento a magazzino, con tasso  $h$  per unità e per unità di tempo;
- un costo di ordinazione, composto da una parte variabile con coefficiente  $c$  e da un costo fisso  $\varphi$ , sostenuto solo quando si ordina una quantità positiva.

**Funzione indicatrice del costo fisso** Per rappresentare il costo fisso si introduce la funzione:

$$\delta(x) = \begin{cases} 1 & \text{se } x > 0, \\ 0 & \text{se } x = 0. \end{cases}$$

**Variabili di stato e controllo** Il livello di inventario  $I_t$  è una variabile di stato, la quantità ordinata  $x_t$  è una variabile di controllo, e la domanda  $d_t$  può essere interpretata come input esogeno nel caso deterministico o come fattore di rischio nel caso stocastico.

**Equazione di transizione** L'equazione di stato è data da (13);  $I_0$  è lo stato iniziale e  $I_T$  è lo stato terminale alla fine dell'orizzonte di pianificazione.

## 7.13 Domanda incerta e modelli alternativi

**Vendite perse** Nel caso di vendite perse, l'equazione di transizione diventa:

$$I_{t+1} = \max\{0, I_t + x_t - d_{t+1}\}.$$

Si introduce inoltre una penalità  $q$  per ogni unità di domanda insoddisfatta, aggiungendo il termine:

$$\sum_{t=1}^T q \max\{0, d_t - (I_{t-1} + x_{t-1})\}.$$

**Clienti pazienti e backlog** Se i clienti sono pazienti, la domanda può essere accumulata come backlog, con una penalità  $b$ , con  $b > h$ .

**Backlog con una variabile di stato** Una prima opzione è mantenere una singola variabile di stato  $I_t$ , rilassandone la non negatività e interpretando valori negativi come backlog.

**Backlog con due variabili di stato** Una seconda opzione è introdurre due variabili di stato:

- inventario disponibile  $O_t \geq 0$ ;
- backlog  $B_t \geq 0$ .

Il costo diventa:

$$\sum_{t=1}^T (hO_t + bB_t),$$

e le equazioni di transizione sono:

$$O_{t+1} = \max\{0, O_t - B_t + x_t - d_{t+1}\},$$

$$B_{t+1} = \max\{0, -O_t + B_t - x_t + d_{t+1}\}.$$

**Ulteriori problematiche** In applicazioni reali possono sorgere ulteriori difficoltà:

- valutazione dello stato terminale;
- stati parzialmente osservabili;
- domanda censurata;
- ritardi dovuti a lead time di consegna;
- incertezza dipendente dallo stato e dalle decisioni.

## 7.14 Uno sguardo al principio di programmazione dinamica: il problema del cammino minimo

**Rete diretta aciclica** Si consideri un problema deterministico: il cammino minimo su una rete diretta aciclica. Una rete diretta è costituita da un insieme di nodi  $N = \{0, 1, 2, \dots, N\}$  e da un insieme di archi  $A$ .

**Archi e costi** Un arco è una coppia ordinata  $(i, j) \in A$ , con  $i, j \in N$ , ed è associato a un costo  $c_{ij} > 0$ .

**Successori e predecessori** Per un nodo  $i \in N$  si definiscono:

$$S_i = \{j \in N \mid (i, j) \in A\}, \quad B_i = \{j \in N \mid (j, i) \in A\}.$$

**Cammini e lunghezza** Un cammino diretto da  $i$  a  $j$  è una sequenza di nodi  $(k_1, k_2, \dots, k_m)$  con  $k_1 = i$ ,  $k_m = j$ , e archi consecutivi in  $A$ . La lunghezza del cammino è la somma dei costi sugli archi.

**Interpretazione dinamica** Ogni nodo è uno stato e ogni arco una transizione possibile. In ogni stato occorre scegliere una transizione in modo da raggiungere lo stato terminale al costo minimo.

**Approccio greedy** Una regola greedy consisterebbe nel muoversi verso il nodo più vicino:

$$\min_{j \in S_i} c_{ij}. \tag{14}$$

Tale approccio non è in generale ottimale.

**Valore degli stati** Si introduce il valore  $V_j$  associato a ciascun nodo successore, risolvendo in  $i$  il problema:

$$\min_{j \in S_i} (c_{ij} + V_j).$$

**Definizione del valore ottimo** Sia  $V_i$  la lunghezza del cammino minimo da  $i$  al nodo terminale  $N$ :

$$V_i := L(i^* \rightarrow N).$$

**Proprietà di annidamento** Se  $j$  appartiene a un cammino ottimo da  $i$  a  $N$ , allora il sottocammino da  $j$  a  $N$  è ottimo.

**Ricorsione di Bellman** Si ottiene quindi l'equazione funzionale:

$$V_i = \min_{j \in S_i} \{c_{ij} + V_j\}, \quad \forall i \in N, \tag{17}$$

con condizione terminale  $V_N = 0$ .

**Programmazione dinamica all'indietro** L'equazione viene risolta etichettando i nodi a partire da quello terminale, seguendo un ordinamento topologico.

**Funzione valore** L'equazione (17) definisce una funzione valore  $V(\cdot) : N \rightarrow \mathbb{R}$  che associa a ciascuno stato il suo valore ottimo.

## 7.15 Esempio numerico

**Condizione terminale e primi nodi etichettabili** Si ha la condizione terminale  $V_7 = 0$  per il nodo terminale, e si considerano i suoi predecessori immediati 4 e 6 (non si può ancora etichettare il nodo 5; si noti che i nodi non sono stati numerati usando un ordinamento topologico). Si trova:

$$V_4 = c_{47} + V_7 = 10 + 0 = 10; \quad V_6 = c_{67} + V_7 = 8 + 0 = 8.$$

Se si denota con  $a_i^*$  il nodo successore selezionato quando si è nel nodo  $i \in N$  (l'azione ottima nello stato  $i$ ), si ha banalmente:

$$a_4^* = 7; \quad a_6^* = 7.$$

**Etichettatura del nodo 5** Ora si può etichettare il nodo 5:

$$V_5 = \min \left\{ \begin{array}{l} c_{56} + V_6 \\ c_{57} + V_7 \end{array} \right\} = \min \left\{ \begin{array}{l} 1 + 8 \\ 5 + 0 \end{array} \right\} = 5 \Rightarrow a_5^* = 7.$$

**Etichettatura del nodo 3** Si considera poi il nodo 3, i cui successori immediati 4, 5 e 6 sono già stati etichettati:

$$V_3 = \min \left\{ \begin{array}{l} c_{34} + V_4 \\ c_{35} + V_5 \\ c_{36} + V_6 \end{array} \right\} = \min \left\{ \begin{array}{l} 3 + 10 \\ 2 + 5 \\ 1 + 8 \end{array} \right\} = 7 \Rightarrow a_3^* = 5.$$

**Etichettatura dei nodi 1, 2 e 0** Allo stesso modo, si ha:

$$\begin{aligned} V_1 &= \min \left\{ \begin{array}{l} c_{13} + V_3 \\ c_{14} + V_4 \end{array} \right\} = \min \left\{ \begin{array}{l} 2 + 7 \\ 1 + 10 \end{array} \right\} = 9 \Rightarrow a_1^* = 3; \\ V_2 &= \min \left\{ \begin{array}{l} c_{23} + V_3 \\ c_{25} + V_5 \end{array} \right\} = \min \left\{ \begin{array}{l} 4 + 7 \\ 7 + 5 \end{array} \right\} = 11 \Rightarrow a_2^* = 3; \\ V_0 &= \min \left\{ \begin{array}{l} c_{01} + V_1 \\ c_{02} + V_2 \end{array} \right\} = \min \left\{ \begin{array}{l} 7 + 9 \\ 6 + 11 \end{array} \right\} = 16 \Rightarrow a_0^* = 1. \end{aligned}$$

**Ricostruzione del cammino minimo** Le azioni ottime in ciascuno stato possono essere memorizzate in una tabella, e si può trovare facilmente il cammino minimo applicando la politica decisionale:  $(0, 1, 3, 5, 7)$ , con lunghezza totale pari a 16.

## 7.16 Cammini minimi su reti strutturate

**Caso strutturato e interpretazione dinamica** Il problema del cammino minimo considerato nella sezione precedente riguarda una rete non strutturata. Un caso più strutturato si verifica quando i nodi della rete corrispondono a stati di un sistema dinamico che evolve nel tempo.

**Contributo terminale e nodo fittizio** Se lo stato terminale  $s_T$  è fissato, non ha senso associare un valore terminale ad esso [Fig. (a)]. Se lo stato terminale è libero, può invece avere senso associare un contributo terminale  $F_T(s_T)$ . In Fig. (b) si aggiunge uno strato temporale al grafo e si introduce un nodo terminale fittizio  $\Omega$ .

## 7.17 Il principio di decomposizione della programmazione dinamica

**Decomposizione in problemi a singolo stadio** Il problema del cammino minimo suggerisce che un problema decisionale multistadio può essere decomposto in una sequenza di problemi più semplici a singolo stadio.

**Problema deterministico** In un problema deterministico, si cerca una sequenza di vettori  $(x_0, x_1, \dots, x_{T-1})$  per ottimizzare una misura di prestazione:

$$\text{opt } H(x_0, x_1, \dots, x_{T-1}; s_0, s_1, \dots, s_T),$$

soggetta a un insieme di vincoli su stati e decisioni.

**Problema stocastico** Nel caso stocastico, il problema è:

$$\text{opt } \mathbb{E}[H(x_0, x_1, \dots, x_{T-1}; s_0, s_1, \dots, s_T)],$$

dove l'aspettazione è presa rispetto a una sequenza di variabili aleatorie  $(\xi_1, \dots, \xi_T)$ . Questa notazione nasconde la vera natura multistadio del problema, poiché ora si dovrebbero trovare una sequenza di funzioni.

**Decisioni come funzioni della storia** Infatti, a parte la decisione iniziale  $x_0$  da prendere qui e ora, i vettori decisionali possono essere funzioni delle decisioni passate e delle realizzazioni osservate dei fattori di rischio:

$$x_t = \mu_t(x[t-1], \xi[t]), \quad t = 1, \dots, T-1. \quad (18)$$

**Riduzione tramite stato** Se si individua un insieme adatto di variabili di stato, si può semplificare (18) e cercare funzioni più semplici che mappano lo stato corrente nella decisione ottima,  $x_t = \mu_t(s_t)$ .

**Additività della funzione obiettivo** Oltre alla struttura markoviana della dinamica, una decomposizione “pulita” è possibile se si assume che la funzione obiettivo abbia forma additiva:

$$\mathbb{E}_0 \left[ \sum_{t=0}^{T-1} \gamma^t f_t(s_t, x_t) + \gamma^T F_T(s_T) \right].$$

**Regola myopica** Una regola decisionale rapida e approssimativa è: quando si è nello stato  $s_t$ , risolvere il problema miopico

$$\text{opt}_{x_t \in X(s_t)} f_t(s_t, x_t),$$

dove  $X(s_t)$  è l'insieme delle decisioni ammissibili nello stato  $s_t$ . È già noto che tale approccio greedy non è atteso funzionare bene in generale: occorre bilanciare obiettivi di breve e lungo periodo introducendo una funzione valore  $V_t(\cdot)$ .

**Significato della funzione valore** L'idea fondamentale della programmazione dinamica è che si può trovare una funzione valore tale da ottenere la prestazione ottima. Il valore  $V_t(s)$  dovrebbe essere il costo/ricavo atteso ottenuto applicando una politica ottima dal tempo  $t$  in avanti, partendo dallo stato  $s$ .

## 7.18 Equazione di Bellman

**Problema a singolo stadio parametrizzato dallo stato** Formalmente, nello stato  $s_t$  al tempo  $t$  si deve risolvere:

$$V_t(s_t) = \text{opt}_{x_t \in X(s_t)} \{ f_t(s_t, x_t) + \gamma \mathbb{E}[V_{t+1}(g_{t+1}(s_t, x_t, \xi_{t+1})) \mid s_t, x_t] \}. \quad (19)$$

Questa equazione funzionale ricorsiva è detta **equazione di Bellman**.

**Politica ottima in feedback** La decisione ottima  $x_t^*$  si ottiene risolvendo un problema di ottimizzazione parametrizzato dallo stato corrente  $s_t$ , usando la conoscenza della funzione valore  $V_{t+1}(\cdot)$ . Nel caso di uno spazio degli stati finito e piccolo, la politica ottima può essere rappresentata in forma tabellare. In generale, la politica è implicita nella sequenza delle funzioni valore. In ogni caso, concettualmente si trova una politica ottima in forma di feedback:

$$x_t^* = \mu_t^*(s_t) \in X(s_t). \quad (20)$$

Raccogliendo tutte le funzioni  $\mu_t^*(\cdot)$  si ottiene la politica complessiva:

$$\mu^* \equiv (\mu_0^*(\cdot), \mu_1^*(\cdot), \dots, \mu_{T-1}^*(\cdot)). \quad (21)$$

Quando si considera una singola funzione  $\mu^*(\cdot)$  indipendente dal tempo si parla di **politica stazionaria**, adatta ai problemi a orizzonte infinito. Talvolta ci si accontenta di una politica subottima, ad esempio ottenuta da un'approssimazione della funzione valore ottima.

## 7.19 Teorema: principio di ottimalità della programmazione dinamica

**Enunciato** Si consideri una politica ottima  $(\mu_0^*(\cdot), \mu_1^*(\cdot), \dots, \mu_{T-1}^*(\cdot))$  per il problema multistadio

$$\text{opt } \mathbb{E}_0 \left[ \sum_{t=0}^{T-1} \gamma^t f_t(s_t, x_t) + \gamma^T F_T(s_T) \right].$$

Si assuma che al tempo  $\tau$  ci si trovi nello stato  $s_\tau$  e si consideri il problema di coda:

$$\text{opt } \mathbb{E}_\tau \left[ \sum_{t=\tau}^{T-1} \gamma^{t-\tau} f_t(s_t, x_t) + \gamma^{T-\tau} F_T(s_T) \right].$$

Allora la politica troncata  $(\mu_\tau^*(\cdot), \mu_{\tau+1}^*(\cdot), \dots, \mu_{T-1}^*(\cdot))$  è ottima per il problema di coda.

**Osservazione sulla dimostrazione** Le dimostrazioni si basano sull'induzione matematica e possono essere piuttosto complicate quando si considerano questioni matematiche sottili.

## 7.20 DP stocastica per orizzonti finiti

**Risoluzione all'indietro** L'equazione (19) è un'equazione di ottimalità e richiede di determinare la funzione valore per ogni istante temporale. Il processo naturale procede all'indietro nel tempo, partendo dalla condizione terminale:

$$V_T(s_T) = F_T(s_T) \quad \forall s_T,$$

cioè il valore dello stato terminale è dato dalla funzione  $F_T(\cdot)$ .

**Ultimo istante decisionale** All'ultimo istante decisionale,  $t = T - 1$ , si risolve per ogni possibile stato  $s_{T-1}$ :

$$V_{T-1}(s_{T-1}) = \text{opt}_{x_{T-1} \in X(s_{T-1})} \{ f_{T-1}(s_{T-1}, x_{T-1}) + \gamma \mathbb{E}[V_T(g_T(s_{T-1}, x_{T-1}, \xi_T)) \mid s_{T-1}, x_{T-1}] \}. \quad (22)$$

Questo è un problema statico ma non miopico, poiché  $V_T(\cdot)$  incorpora l'effetto della decisione  $x_{T-1}$  sullo stato terminale.

**Ricorsione fino a  $V_1$**  Risolvendo (22) per ogni stato  $s_{T-1}$  si costruisce la funzione valore  $V_{T-1}(\cdot)$ . Poi, svolgendo la ricorsione all'indietro nel tempo, si ottengono  $V_{T-2}(s_{T-2})$  e così via, fino a  $V_1(s_1)$ .

**Prima decisione** Infine, dato lo stato iniziale  $s_0$ , la prima decisione ottima si trova risolvendo:

$$V_0(s_0) = \text{opt}_{x_0 \in X(s_0)} \left\{ f_0(s_0, x_0) + \gamma \mathbb{E}[V_1(g_1(s_0, x_0, \xi_1)) \mid s_0, x_0] \right\}. \quad (23)$$

**Uso delle funzioni valore: caso deterministico** Come sfruttare la conoscenza delle funzioni valore  $V_t(\cdot)$ ? In un contesto deterministico, si può trovare la sequenza di decisioni ottime  $x_t^*$  risolvendo una sequenza di problemi a singolo stadio e aggiornando lo stato secondo le decisioni applicate.

**Uso delle funzioni valore: simulazione Monte Carlo** In un contesto stocastico, si può eseguire una simulazione Monte Carlo come segue:

- dato lo stato iniziale  $s_0$  e la funzione valore  $V_1(\cdot)$ , si risolve il problema del primo stadio e si trova  $x_0^*$ ;
- si campiona  $\xi_1$  e si genera lo stato successivo  $s_1 = g_1(s_0, x_0^*, \xi_1)$ ;
- dato lo stato  $s_1$  e la funzione valore  $V_2(\cdot)$ , si risolve il problema del secondo stadio e si trova  $x_1^*$ ;
- si ripete il processo fino a generare l'ultima decisione  $x_{T-1}^*$  e lo stato terminale  $s_T$ .

## 7.21 DP stocastica per orizzonti infiniti

**Problema a orizzonte infinito** La forma ricorsiva (19) deve essere adattata per un problema a orizzonte infinito del tipo:

$$\text{opt } \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t f(s_t, x_t) \right], \quad (24)$$

dove si assume che i costi immediati siano limitati e  $\gamma < 1$ , così che la serie converga a un valore finito. Si elimina l'indice  $t$  dal contributo immediato e dalla funzione di transizione:

$$s_{t+1} = g(s_t, x_t, \xi_{t+1}).$$

**Equazione al punto fisso** L'equazione funzionale diventa:

$$V(s) = \text{opt}_{x \in X(s)} \left\{ f(s, x) + \gamma \mathbb{E}[V(g(s, x, \xi))] \right\}, \quad (25)$$

dove  $X(s)$  è l'insieme delle decisioni ammissibili nello stato  $s$ . Ora la funzione valore è definita come **punto fisso** di un operatore potenzialmente complicato, poiché  $V(s)$  compare su entrambi i lati dell'equazione. È necessario un metodo iterativo per risolvere (25).

# 8 CAP 8 – Implementazione della programmazione dinamica

## 8.1 Allocazione discreta di risorse: il problema dello zaino

**Definizione del problema** Il problema dello **zaino** richiede di selezionare un sottoinsieme di oggetti di **valore totale massimo**, rispettando un vincolo di **capacità** (budget).

Per modellare la selezione di ciascun oggetto si introducono variabili decisionali **binarie**:

$$x_k = \begin{cases} 1 & \text{se l'oggetto } k \text{ è selezionato,} \\ 0 & \text{altrimenti.} \end{cases}$$

Il problema può essere formulato come un problema di **programmazione lineare binaria pura**:

$$\max \sum_{k=1}^n v_k x_k$$

$$\begin{aligned} \text{s.t. } & \sum_{k=1}^n w_k x_k \leq B \\ & x_k \in \{0, 1\} \quad \forall k. \end{aligned}$$

Si assume un'allocazione discreta del budget, poiché la selezione di un'attività è una decisione **tutto-o-niente**.

**Motivazione dell'approccio DP** Il problema può essere risolto in modo efficiente con algoritmi **branch-and-cut** per la programmazione lineare intera, ma qui si adotta un approccio basato sul principio della **programmazione dinamica**.

## 8.2 Riformulazione sequenziale del problema

**Indice temporale fittizio e stadi decisionali** Il problema non è intrinsecamente dinamico, ma può essere riformulato come un problema di **allocazione sequenziale di risorse** introducendo un indice discreto fittizio  $k$ , associato agli oggetti.

Per ogni stadio  $k = 1, \dots, n$ , si decide se includere o meno l'oggetto  $k$  nel sottoinsieme.

**Stato e dipendenza dalle decisioni passate** Allo stadio  $k = 1$  si dispone del budget  $B$  e si deve scegliere se includere il primo oggetto, affrontando un compromesso tra **ricompensa immediata** e **ricompense future**. Allo stadio  $k = n$  il problema è banale, poiché resta da valutare solo l'ultimo oggetto dato il budget residuo.

La selezione degli oggetti successivi dipende dalle decisioni passate soltanto attraverso il **budget residuo**. La variabile di stato naturale allo stadio  $k$  è il budget disponibile  $s_k$  prima di selezionare l'oggetto  $k$ , mentre la variabile decisionale è la variabile binaria  $x_k$ .

**Equazione di transizione dello stato** Poiché non esiste l'oggetto  $k = 0$ , si usa la transizione:

$$s_{k+1} = s_k - w_k x_k, \quad k = 1, \dots, n,$$

con condizione iniziale

$$s_1 = B.$$

Assumendo che i pesi  $w_k$  siano **intesti**, anche la variabile di stato assume valori **intesti**.

## 8.3 Funzione valore e ricorsione DP

**Definizione della funzione valore** Si definisce la funzione valore:

$$V_k(s) := \text{profitto della selezione ottimale degli oggetti } \{k, k+1, \dots, n\} \text{ con budget residuo } s.$$

Nel caso discreto, è possibile **tabulare** tutte le funzioni  $V_k(s)$  per  $k = 1, \dots, n$  e  $s \in \{0, \dots, B\}$ .

**Equazione di Bellman** La ricorsione DP è:

$$V_k(s) = \begin{cases} V_{k+1}(s) & \text{se } 0 \leq s < w_k, \\ \max\{V_{k+1}(s), V_{k+1}(s - w_k) + v_k\} & \text{se } w_k \leq s \leq B. \end{cases}$$

**Interpretazione della ricorsione** Allo stadio  $k$ , si considera l'oggetto  $k$ . Se  $s < w_k$ , l'oggetto non entra nel budget residuo e quindi la funzione valore resta **invariata**. Se  $s \geq w_k$ , si confrontano due alternative: includere l'oggetto (ottenendo  $v_k$  e aggiornando lo stato a  $s - w_k$ ) oppure escluderlo (lasciando lo stato a  $s$ ). Poiché la decisione è **binaria**, l'ottimizzazione a singolo stadio è immediata.

**Condizione terminale** Per l'ultimo oggetto  $k = n$ :

$$V_n(s) = \begin{cases} 0 & \text{se } 0 \leq s < w_n, \\ v_n & \text{se } w_n \leq s \leq B. \end{cases}$$

## 8.4 Implementazione MATLAB del problema dello zaino

**Obiettivo della funzione** La funzione `DPKnapsack` ha lo scopo di calcolare la soluzione ottima del problema dello **zaino** dato un vettore di **valori**, un vettore di **pesi** e una **capacità**. L'output è la selezione ottima come vettore **binario** e il **valore totale** ottenuto.

**Idea di memorizzazione: tabelle di valore e decisione** La funzione `valore` viene memorizzata in una tabella `valueTable`, indicizzata per **stadio** (oggetto) e per **stato** (budget residuo). Poiché MATLAB indica a partire da 1, lo stato  $s$  corrisponde alla riga  $s + 1$ .

In parallelo, la tabella `decisionTable` memorizza, per ogni coppia **stato–stadio**, la **decisione ottima** (includere o non includere).

**Ricorsione backward e ricostruzione forward** La tabulazione implementa una ricorsione **backward** sugli stadi, calcolando le funzioni valore a ritroso. Successivamente, la soluzione ottima viene ricostruita procedendo **forward** sugli oggetti, usando la tabella delle decisioni e aggiornando lo stato (budget residuo) coerentemente.

**Complessità computazionale** Assumendo dati interi, la complessità è  $O(nB)$ , quindi **pseudo-polynomiale**. Quando  $B$  è grande, occorre costruire una tabella molto grande e l'algoritmo può diventare poco efficiente.

## 8.5 Allocazione continua del budget

**Formulazione del problema** Si considera una versione **continua** del problema di allocazione delle risorse: un budget  $B$  deve essere allocato tra  $n$  attività, con decisioni continue  $x_k \geq 0$ ,  $k = 1, \dots, n$ . Il contributo di profitto dell'attività  $k$  dipende dall'allocazione tramite una funzione  $f_k(\cdot)$  **crescente e concava**:

$$\begin{aligned} & \max \sum_{k=1}^n f_k(x_k) \\ \text{s.t. } & \sum_{k=1}^n x_k \leq B, \quad x_k \geq 0 \quad \forall k. \end{aligned}$$

**Uso completo del budget** Poiché le funzioni di profitto sono **strettamente crescenti**, si può assumere che il budget sia interamente utilizzato, quindi il vincolo di budget è soddisfatto come **uguaglianza** nella soluzione ottima.

**Soluzione interna e Lagrangiana** Assumendo una soluzione interna  $x_k^* > 0$ , il problema diventa un problema non lineare con un solo vincolo di uguaglianza. Introducendo il moltiplicatore di Lagrange  $\lambda$ , la lagrangiana è:

$$L(x, \lambda) = \sum_{k=1}^n \sqrt{x_k} + \lambda \left( \sum_{k=1}^n x_k - B \right).$$

Le condizioni del primo ordine sono:

$$\frac{1}{2\sqrt{x_k}} + \lambda = 0, \quad k = 1, \dots, n,$$

$$\sum_{k=1}^n x_k - B = 0,$$

che implicano un'allocazione **uniforme**:

$$x_k^* = \frac{B}{n}, \quad k = 1, \dots, n.$$

In questo caso, la funzione obiettivo è **concava** e le condizioni di ottimalità sono **necessarie e sufficienti**.

## 8.6 Allocazione continua del budget: formulazione DP

**Riformulazione DP** Il problema può essere riformulato in un quadro di **programmazione dinamica** introducendo un indice discreto fittizio  $k = 1, \dots, n$  associato alle attività. Sia  $V_k(s)$  il profitto ottimale ottenibile allocando un budget residuo  $s$  alle attività  $\{k, k+1, \dots, n\}$ .

**Transizione dello stato** La transizione dello stato è:

$$s_{k+1} = s_k - x_k,$$

con condizione iniziale  $s_1 = B$ .

**Equazioni di ottimalità** Le funzioni valore soddisfano:

$$V_k(s_k) = \max_{0 \leq x_k \leq s_k} \{f_k(x_k) + V_{k+1}(s_k - x_k)\},$$

con condizione terminale:

$$V_n(s_n) = \max_{0 \leq x_n \leq s_n} f_n(x_n) = f_n(s_n).$$

Il vincolo  $s_k \geq 0$  è garantito imponendo il vincolo sulla decisione  $0 \leq x_k \leq s_k$ , poiché la transizione è deterministica.

**Funzioni valore in spazio infinito-dimensionale** Nel caso continuo, la funzione valore  $V_k(s)$  è un oggetto in uno spazio funzionale **infinito-dimensionale**. Poiché la valutazione è possibile solo su un insieme finito di stati, occorre usare **approssimazione** o **interpolazione** per stimare i valori fuori dalla griglia.

## 8.7 Intermezzo: interpolazione con spline cubiche in MATLAB

**Limiti dell'interpolazione polinomiale** È noto che l'interpolazione polinomiale può soffrire di **oscillazioni inaccettabili**. Una funzione **concava e monotona crescente** può essere approssimata da una funzione **non monotona**, con effetti negativi sulle procedure di ottimizzazione.

**Spline cubiche come alternativa** Un'alternativa standard consiste nell'uso di funzioni polinomiali a tratti di ordine più basso, in cui ogni tratto è associato a un sottointervallo della griglia. Una scelta comune sono le **spline cubiche**, ottenute in MATLAB tramite **spline** (costruzione) e **ppval** (valutazione in punti arbitrari, anche fuori griglia).

**Accuratezza, griglia e questioni aperte** L'errore può essere significativo con griglie grossolane e migliora con griglie più fitte. Restano aperti temi come: la garanzia che una funzione monotona venga approssimata da una spline monotona, la scelta dei nodi per un buon compromesso tra costo computazionale ed errore, e la generalizzazione a dimensioni superiori.

## 8.8 Risoluzione numerica del problema continuo tramite DP

**Idea della discretizzazione e interpolazione** Si usa una **spline cubica** per approssimare la funzione valore del problema continuo. Si imposta una griglia uniforme su  $[0, B]$ , replicata per ciascuno stadio. La griglia contiene  $m + 1$  valori di stato, con passo  $\delta s = B/m$ , cioè stati del tipo  $j \delta s$ ,  $j = 0, 1, \dots, m$ .

**Risoluzione dei sottoproblemi e policy implicita** Per ogni punto della griglia si risolve un sottoproblema del tipo:

$$V_k(s_k) = \max_{0 \leq x_k \leq s_k} \{f_k(x_k) + V_{k+1}(s_k - x_k)\},$$

tramite ottimizzazione numerica. I valori fuori dalla griglia sono stimati via spline cubiche. La condizione al bordo su  $V_n(\cdot)$  è banale e si arresta il calcolo a  $V_2(\cdot)$  quando si vuole risolvere il problema per un budget specifico  $B$ . La politica ottima  $x_t^* = \mu_t^*(s_t)$  non è memorizzata in una tabella esplicita, ma è **implicita** nella sequenza delle funzioni valore.

## 8.9 Implementazione MATLAB: idea generale

**Scopo di `findPolicy`** La funzione `findPolicy` serve a costruire, stadio per stadio, un'**approssimazione** delle funzioni valore tramite **spline cubiche**, usando una griglia uniforme del budget. Riceve il budget totale, una lista di funzioni di profitto (una per attività) e il numero di punti della griglia. Restituisce una lista di spline che rappresentano le funzioni valore approssimate; la prima funzione non viene usata direttamente quando l'interesse è risolvere il problema per un budget iniziale specifico.

**Scopo di `applyPolicy`** La funzione `applyPolicy` applica la politica ottima **in avanti nel tempo** utilizzando le funzioni valore approssimate: a ogni stadio sceglie l'allocazione che massimizza la somma tra contributo immediato e valore futuro, aggiorna il budget residuo e costruisce il vettore delle allocazioni ottime insieme al valore complessivo ottenuto.

## 8.10 Controllo stocastico delle scorte

**Impostazione del problema** Si considera una variazione **stocastica** del problema di lot-sizing, assumendo una **domanda aleatoria discreta**. In questo caso è naturale adottare una rappresentazione **tabellare** della funzione valore.

**Dinamica dello stato con stockout (lost sales)** Occorre specificare la dinamica dello stato in caso di stockout. Qui si assumono **vendite perse**:

$$I_{t+1} = \max\{0, I_t + x_t - d_{t+1}\}$$

dove  $(d_t)_{t=1,\dots,T}$  è una sequenza di variabili i.i.d., e  $x_t$  è la quantità ordinata al tempo  $t$  e consegnata immediatamente.

**Sequenza degli eventi** Al tempo  $t$  si osserva l'inventario a mano  $I_t$ . Poi si decide l'ordine  $x_t$ , che viene ricevuto immediatamente e porta la disponibilità a  $I_t + x_t$ . Durante l'intervallo  $t+1$  si osserva la domanda aleatoria  $d_{t+1}$  e si aggiorna l'inventario secondo la dinamica sopra.

**Spazio degli stati e azioni ammissibili** Per tabulare la funzione valore occorre fissare un limite superiore allo stato. Si assume un vincolo  $I_t \leq I_{\max}$ , che implica:

$$X(I_t) = \{0, 1, \dots, I_{\max} - I_t\}.$$

**Costo immediato e stato terminale** Il costo immediato include un costo lineare d'ordine  $c x_t$  e una penalità quadratica sull'inventario contabile dopo la domanda del periodo successivo:

$$\boxed{\phi(I_t + x_t - d_{t+1})^2.}$$

L'inventario fisico non può essere negativo ed è determinato dalla dinamica con  $\max\{0, \cdot\}$ ; un inventario contabile negativo rappresenta domanda non soddisfatta. Per semplicità, il costo terminale è nullo:

$$F_T(I_T) = 0.$$

**Ricorsione DP con costo immediato stocastico** Il costo immediato dipende dalla realizzazione del fattore aleatorio nel periodo successivo alla decisione; quindi nella ricorsione compare un termine **stocastico**. La ricorsione è:

$$\boxed{V_t(I_t) = \min_{x_t \in X(I_t)} \mathbb{E}_{d_{t+1}} \left[ c x_t + \phi(I_t + x_t - d_{t+1})^2 + V_{t+1}(\max\{0, I_t + x_t - d_{t+1}\}) \right]}$$

per  $t = 0, 1, \dots, T-1$  e  $I_t \in \{0, 1, 2, \dots, I_{\max}\}$ .

**Modellare l'incertezza** Poiché la domanda è discreta e i.i.d., è sufficiente una **funzione di massa di probabilità**, cioè un vettore  $\pi_k$  per i possibili valori  $k = 0, 1, \dots, d_{\max}$ . Occorre inoltre specificare lo stato iniziale  $I_0$  e l'orizzonte  $T$ .

## 8.11 Lot-sizing stocastico: idea delle procedure MATLAB

**Scopo di MakePolicy** La funzione `MakePolicy` serve ad apprendere le funzioni valore **a ritroso (backward)** e la **politica ottima** in forma tabellare, dato un limite massimo di inventario, una distribuzione discreta della domanda e i parametri economici. L'output è una tabella dei valori (funzione valore per ogni stato e istante) e una tabella delle azioni (decisione ottima d'ordine per ogni stato e istante).

**Scopo di SimulatePolicy** La funzione `SimulatePolicy` serve a **simulare** l'applicazione della politica ottima su molti scenari di domanda generati casualmente, per ottenere una stima statistica del costo totale e confrontarla con il valore previsto dalla funzione valore.

## 8.12 Sfruttare la struttura: cammini minimi per il lot-sizing deterministico

**Motivazione** Nel problema toy di lot-sizing è semplice memorizzare funzioni valore e politica in una tabella, ma ciò è **inefficiente** quando i valori possibili di domanda sono molti, e diventa impossibile con spazio degli stati **continuo**. Talvolta è possibile semplificare drasticamente il problema sfruttando la **struttura**.

**Lot-sizing deterministico con costi fissi e holding** Si consideri una versione deterministica con soli **costi fissi d'ordine  $\phi$**  e **costi di giacenza  $h$** , con inventario iniziale e finale pari a zero, e domanda non nulla nel primo periodo. In linea di principio, una ricorsione DP è:

$$\boxed{V_t(I_t) = \min_{x_t \geq d_{t+1} - I_t} \{\phi \delta(x_t) + h(I_t + x_t - d_{t+1}) + V_{t+1}(I_{t+1})\}, \quad t = 0, \dots, T-1}$$

con condizione:

$$\boxed{V_T(I_T) \equiv 0.}$$

Il vincolo su  $x_t$  garantisce che la domanda sia sempre soddisfatta e che lo stato non diventi negativo.

**Teorema: proprietà di Wagner–Whitin** Per il lot-sizing deterministico non capacitato con **costi fissi** e **costi lineari di inventario**, esiste una soluzione ottima tale che:

$$I_t x_t = 0, \quad t = 0, 1, \dots, T-1.$$

Il messaggio è che non è mai ottimale ordinare quando l'inventario è positivo: si ordina solo se l'inventario è **nullo**.

**Rete di flusso e interpretazione** Considerando una rappresentazione come rete, deve valere il bilancio globale:

$$\sum_{t=0}^{T-1} x_t = \sum_{t=1}^T d_t.$$

Si introduce un nodo fittizio 0 il cui inflow rappresenta l'ammontare totale ordinato. I nodi  $t = 1, \dots, T$  rappresentano gli istanti associati alla fine di ciascun intervallo, cioè la domanda si soddisfa alla fine dell'intervallo.

Il bilancio al nodo  $t$  corrisponde alla transizione:

$$I_t = I_{t-1} + x_{t-1} - d_t.$$

Se, contrariamente al teorema,  $I_{t-1} > 0$  e  $x_{t-1} > 0$ , allora reindirizzando l'inflow orizzontale  $I_{t-1}$  lungo l'arco d'ordine associato a  $x_{t-1}$  si può migliorare il costo totale.

**Conseguenza pratica: scelte d'ordine candidate** All'istante  $t$  basta considerare solo:

$$x_t \in \left\{ 0, d_{t+1}, (d_{t+1} + d_{t+2}), (d_{t+1} + d_{t+2} + d_{t+3}), \dots, \sum_{\tau=t+1}^T d_{\tau} \right\}.$$

**Riformulazione come cammino minimo** Con questa proprietà, il problema a singolo item si riformula come un **cammino minimo** su una rete piccola: dal nodo iniziale 0 al nodo terminale, scegliendo archi che rappresentano quanti periodi vengono coperti dal prossimo ordine. Il costo di ciascun arco incorpora il **costo fisso d'ordine** e il **costo di giacenza** risultante. Poiché la rete ha un numero limitato di nodi, questa formulazione porta a un algoritmo molto efficiente di complessità **polinomiale**.

## 8.13 Lot-sizing stocastico: politiche $S$ e $(s, S)$

**Backlog e penalità convessa** La proprietà di Wagner–Whitin non vale nel caso stocastico, ma in alcuni casi esistono risultati strutturali. Si assume che non vi siano **vendite perse** e che sia consentito backlog, con costo totale che include:

$$q(s) = h \max\{0, s\} + b \max\{0, -s\},$$

dove  $s$  può essere positivo (inventario) o negativo (backlog),  $h$  è il costo di giacenza e  $b > h$  è il costo di backlog. La funzione  $q(\cdot)$  è **convessa** e tende a  $+\infty$  quando  $s \rightarrow \pm\infty$ .

**Costo variabile e obiettivo atteso** Trascurando i costi fissi e includendo un costo variabile lineare unitario  $c$ , l'obiettivo è minimizzare:

$$\mathbb{E}_0 \left[ \sum_{t=0}^{T-1} (c x_t + q(I_t + x_t - d_{t+1})) \right].$$

**Ricorsione DP e funzione  $H(\cdot)$**  La ricorsione DP è:

$$V_t(I_t) = \min_{x_t \geq 0} \{c x_t + H(I_t + x_t) + \mathbb{E}[V_{t+1}(I_t + x_t - d_{t+1})]\},$$

dove:

$$H(y_t) := \mathbb{E}[q(y_t - d_{t+1})] = h \mathbb{E}[\max\{0, y_t - d_{t+1}\}] + b \mathbb{E}[\max\{0, d_{t+1} - y_t\}].$$

Si introduce  $y_t$  come inventario disponibile dopo l'ordine (lead time nullo):

$$y_t := I_t + x_t,$$

con condizione terminale:

$$V_T(I_T) = 0.$$

Si assume distribuzione della domanda costante nel tempo.

**Riscrittura tramite  $G_t(\cdot)$**  La ricorsione si riscrive come:

$$V_t(I_t) = \min_{y_t \geq I_t} G_t(y_t) - c I_t,$$

dove:

$$G_t(y_t) = c y_t + H(y_t) + \mathbb{E}[V_{t+1}(y_t - d_{t+1})].$$

**Convessità e livello obiettivo  $S_t$**  Si assume (senza prova) che  $V_t(\cdot)$  e  $G_t(\cdot)$  siano **convesse** per ogni  $t$ , e che  $G_t(\cdot) \rightarrow +\infty$  quando  $y \rightarrow \pm\infty$ . Ne segue che  $G_t(\cdot)$  ha un minimizzatore non vincolato finito:

$$S_t = \arg \min_{y_t \in \mathbb{R}} G_t(y_t).$$

**Politica base-stock (order-up-to)** La politica ottima è una politica **base-stock**:

$$x_t^* = \mu_t^*(I_t) = \begin{cases} S_t - I_t, & \text{se } I_t < S_t, \\ 0, & \text{se } I_t \geq S_t. \end{cases}$$

I valori  $S_t$  sono livelli obiettivo: si ordina quanto serve per raggiungere il target ottimo. In orizzonte finito, il problema è determinare la sequenza ottima di livelli  $S_t$ .

**Politiche  $(s, S)$  con costi fissi** Se si introducono costi fissi, si perde convessità. Tuttavia una proprietà correlata (K-convessità) porta a una politica ottima:

$$\mu_t^*(I_t) = \begin{cases} S_t - I_t, & \text{se } I_t < s_t, \\ 0, & \text{se } I_t \geq s_t, \end{cases}$$

dipendente da due sequenze  $s_t$  e  $S_t$ , con  $s_t \leq S_t$ . In ambiente stazionario è ottima una politica stazionaria  $(s, S)$ . Si ordina solo quando l'inventario è sotto **small**  $s$ , riportandolo a **big**  $S$ ;  $S - s$  è una quantità minima d'ordine che aiuta a controllare i costi fissi.

## 8.14 Le maledizioni della programmazione dinamica

**Limiti della DP** La DP è potente e flessibile, ma presenta limitazioni importanti.

**Maledizione della dimensionalità dello stato** Serve la funzione valore per ogni elemento dello spazio degli stati. Se lo spazio è finito e piccolo, si possono usare tabelle; per spazi enormi ciò non è fattibile.

**Maledizione dell'ottimizzazione** La DP decomponete un problema multistadio in sottoproblemi a singolo stadio, ma anche questi sottoproblemi possono essere difficili da risolvere.

**Maledizione dell'aspettativa** Se i fattori di rischio  $\xi_t$  sono continui, l'aspettativa richiede integrali multidimensionali difficili; serve quindi una discretizzazione.

**Maledizione della modellazione** Il sistema può essere troppo complesso per avere un modello esplicito delle transizioni di stato. In DP il problema è più complicato perché le transizioni sono almeno in parte influenzate dalle decisioni di controllo.

## 9 CAP 9 – Modelli della programmazione dinamica

## 10 CAP 9 – Modeling for Dynamic Programming

### 10.1 Frontespizio e riferimenti

**Titolo e autore** Business Analytics (2020/21). **Modeling for dynamic programming**. Prof. Paolo Brandimarte, Dipartimento di Scienze Matematiche, Politecnico di Torino.

**Riferimenti** Le slide sono tratte dal libro: **P. Brandimarte**, *From Shortest Paths to Reinforcement Learning: A MATLAB-Based Introduction to Dynamic Programming*, Springer, 2021. Ulteriori riferimenti: P. Brandimarte, *An Introduction to Financial Markets: A Quantitative Approach*, Wiley, 2018; J.Y. Campbell, L.M. Viceira, *Strategic Asset Allocation*, Oxford University Press, 2002; W.B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality* (2nd ed.), Wiley, 2011; G.J. van Ryzin, K.T. Talluri, *An Introduction to Revenue Management*, INFORMS Tutorials in OR, 2005.

### 10.2 Principio di modellazione nella programmazione dinamica

**Idea generale della programmazione dinamica** Il principio della **programmazione dinamica (DP)** è un concetto flessibile per la decomposizione di un **problema decisionale multistadio** in una sequenza di **problemi a singolo stadio**, bilanciando il **contributo immediato** e i **contributi attesi** delle decisioni future.

**Funzione valore e ricorsione DP** Il principio si fonda sulla **funzione valore**, definita dalla ricorsione di DP:

$$V_t(s_t) = \text{opt}_{x_t \in X(s_t)} \{ f_t(s_t, x_t) + \gamma \mathbb{E}[V_{t+1}(s_{t+1}) | s_t, x_t] \}. \quad (1)$$

**Forme alternative della ricorsione DP** L'Eq. (1) è solo una possibile forma della ricorsione di DP. Si possono adottare forme più specifiche, ad esempio quando i **fattori di rischio** sono **variabili aleatorie discrete** e l'aspettazione si riduce a una **somma**.

**Scambio tra aspettazione e ottimizzazione** In alcuni casi si adotta una riformulazione più radicale, in cui si scambiano gli operatori di **aspettazione** e **ottimizzazione**. Ciò può dipendere dalla **struttura informativa del problema** oppure essere il risultato di una manipolazione volta a evitare la soluzione di un **sottoproblema stocastico difficile**.

**Q-learning e stati post-decisione** Un caso ben noto è il **Q-learning**, una forma di **reinforcement learning** in cui  $V(s)$  è sostituita da **Q-factors**  $Q(s, x)$ , che rappresentano il valore delle **coppie stato-azione**. Più in generale, lo scambio tra ottimizzazione e aspettazione può essere ottenuto introducendo il concetto di **stato post-decisione**.

**Scelta della rappresentazione dello stato** Trovare una descrizione adeguata dello **stato del sistema** può richiedere un'attenta modellazione: può essere necessaria una **ridefinizione dello spazio degli stati** per eliminare **dipendenze dal percorso** e ottenere un modello **markoviano aumentato ed equivalente**.

**Natura degli stati e delle decisioni** Sebbene sia naturale pensare a stati e decisioni come **scalar**i o **vettori**, essi possono consistere in oggetti diversi, ad esempio **insiemi**.

**Ruolo delle capacità di modellazione** Le **capacità di modellazione** procedono di pari passo con la **conoscenza algoritmica** nell'affrontare un problema decisionale tramite DP; tali capacità si affinano solo con l'esperienza su una **vasta gamma di problemi**.

### 10.3 Processi decisionali di Markov finiti

**Definizione di MDP** Il termine **processo decisionale di Markov (MDP)** è riservato a problemi con **spazi di stato e spazi delle azioni discreti**. Il termine **azione** è spesso adottato per riferirsi alle decisioni di controllo. Stati e azioni discreti possono essere **enumerati** e associati a numeri interi.

**MDP finiti e notazione degli stati** Si considerano **MDP finiti**. Anche se gli stati possono corrispondere a vettori in uno spazio multidimensionale, si usa una notazione del tipo

$$i, j \in S = \{1, \dots, N\},$$

dove  $N$  è la **dimensione dello spazio degli stati**.

**Azioni ammissibili** Si usa  $a$  o  $a_t$  per indicare le azioni;  $A_t(i)$  o  $A(i)$  denota l'insieme delle **azioni ammissibili** nello stato  $i$  al tempo  $t$ . L'insieme di tutte le azioni possibili è  $A$ .

**Dinamica markoviana controllata** Il sistema sottostante è una **catena di Markov finita a tempo discreto**. Nei MDP le **probabilità di transizione** dipendono dall'azione selezionata:

$$\pi_{t+1}(i, a, j)$$

è la probabilità di transizione dallo stato  $i$  allo stato  $j$  durante l'intervallo  $t + 1$ , dopo aver scelto l'azione  $a$  al tempo  $t$ .

### 10.4 Esempio: controllo stocastico di inventario

**Sintesi dell'esempio** Dall'esempio si evince come, con **spazio di stato finito** e **azioni ammissibili dipendenti dallo stato**, le **matrici di transizione** dipendano dall'azione e possano essere rappresentate in forma tabellare.

### 10.5 Ricorsioni DP per MDP

**Ricorsione a orizzonte finito** Nel contesto MDP, la funzione valore al tempo  $t$  è un vettore finito-dimensionale con componenti  $V_t(i)$  e soddisfa:

$$V_t(i) = \text{opt}_{a \in A(i)} \left\{ f_t(i, a) + \gamma \sum_{j \in S} \pi_{t+1}(i, a, j) V_{t+1}(j) \right\}, \quad i \in S. \quad (2)$$

**Caso a orizzonte infinito scontato** Nel caso a **orizzonte infinito scontato**:

$$V(i) = \text{opt}_{a \in A(i)} \left\{ f(i, a) + \gamma \sum_{j \in S} \pi(i, a, j) V(j) \right\}, \quad i \in S. \quad (3)$$

**Contributo immediato stocastico** Se il contributo immediato è stocastico e dipende dallo stato successivo, lo si denota  $h(i, a, j)$  e si riscrive:

$$V(i) = \text{opt}_{a \in A(i)} \sum_{j \in S} \pi(i, a, j) \{h(i, a, j) + \gamma V(j)\}, \quad i \in S. \quad (4)$$

**Relazione tra contributi** In principio,

$$f(i, a) = \sum_{j \in S} \pi(i, a, j) h(i, a, j),$$

ma ciò può essere **impraticabile** quando  $N$  è finito ma enorme oppure quando le **probabilità di transizione non sono note**.

## 10.6 Esempio: arresto ottimo ricorrente

**Sintesi dell'esempio** Dall'esempio si evince il **trade-off** tra **ricompensa immediata** e **attesa** di opportunità migliori, e come la struttura della catena consenta una formulazione DP esplicita sui valori  $V(k)$ .

## 10.7 Valutazione delle politiche e Q-factors

**Policy evaluation e policy iteration** Una politica stazionaria ammissibile  $\mu$  associa a ciascuno stato  $i$  un'azione  $a = \mu(i) \in A(i)$ . La funzione valore associata a  $\mu$  si ottiene risolvendo un sistema lineare:

$$V^\mu(i) = f(i, \mu(i)) + \gamma \sum_{j \in S} \pi(i, \mu(i), j) V^\mu(j), \quad i \in S. \quad (5)$$

Questa relazione è fondamentale nei metodi basati su **policy iteration**: si valuta una politica candidata e poi si tenta di migliorarla.

**Definizione di Q-factor per una politica** Il **Q-factor** associato a  $\mu$  è la mappa  $S \times A \rightarrow \mathbb{R}$ :

$$Q^\mu(i, a) := f(i, a) + \gamma \sum_{j \in S} \pi(i, a, j) V^\mu(j). \quad (6)$$

L'idea è applicare  $a$  nello stato  $i$  e poi seguire la politica  $\mu$ .

**Q-factors ottimi e relazione con  $V$**  Sostituendo la funzione valore ottima in (6) si ottengono i **Q-factors ottimi**. Vale:

$$V(j) \equiv \text{opt}_{a \in A(j)} Q(j, a), \quad j \in S.$$

**Ricorsione DP in termini di Q-factors** La ricorsione di DP può essere riscritta come:

$$Q(i, a) = f(i, a) + \gamma \sum_{j \in S} \pi(i, a, j) [\text{opt}_{\tilde{a} \in A(j)} Q(j, \tilde{a})], \quad i \in S, a \in A(i). \quad (8)$$

**Vantaggi e svantaggi** Lo **svantaggio** è che si passa da  $V(i)$  a funzioni **stato–azione**  $Q(i, a)$ , aggravando potenzialmente la **curse of dimensionality**. Il **vantaggio** è lo scambio tra **aspettazione** e **ottimizzazione**, utile anche perché i Q-factors possono essere **appresi tramite campionamento statistico** (DP **model-free**) e, se necessario, rappresentati con architetture di approssimazione dalla regressione lineare alle reti neurali profonde.

## 10.8 Diverse forme di DP stocastica

**Ricorsione DP di base** Si consideri nuovamente la ricorsione di DP:

$$V_t(s_t) = \text{opt}_{x_t \in X(s_t)} \{f_t(s_t, x_t) + \gamma \mathbb{E}[V_{t+1}(s_{t+1}) | s_t, x_t]\}.$$

**Ipotesi sottostanti** Si osserva lo **stato**  $s_t$  all’istante  $t$ ; si prende una **decisione ammissibile**  $x_t \in X(s_t)$ ; si osserva un **contributo immediato**  $f_t(s_t, x_t)$ ; si passa a un nuovo stato  $s_{t+1}$  che dipende dai **fattori di rischio realizzati**  $\xi_{t+1}$ , secondo una distribuzione che può dipendere da  $s_t$  e  $x_t$ .

**Difficoltà e motivazione dello scambio** Per determinare  $V_t(\cdot)$  a partire da  $V_{t+1}(\cdot)$  si dovrebbe risolvere un problema di **ottimizzazione stocastica** che può includere un’**aspettazione impegnativa**; nei MDP i **Q-factors** consentono di scambiare **ottimizzazione** e **aspettazione**, e talvolta lo scambio è richiesto dalla **struttura informativa**.

## 10.9 Esempio: lot-sizing con lookahead limitato

**Sintesi dell’esempio** Dall’esempio si evince che, se la domanda del periodo successivo è **nota** al tempo  $t$ , la ricorsione naturale può assumere una forma in cui l’**ottimizzazione interna** è **deterministica** e l’**aspettazione** è **esterna**.

## 10.10 Variabili di stato post-decisione

**Ricorsione in forma scambiata** Una forma alternativa è:

$$V_t(s_t) = \mathbb{E}_t [\text{opt}_{x_t \in X(s_t)} \{f_t(x_t, s_t) + \gamma V_{t+1}(s_{t+1})\}]. \quad (9)$$

**Vantaggi potenziali** Il problema di ottimizzazione interno è **deterministico** e l’aspettazione esterna può essere stimata tramite **campionamento statistico**.

**Introduzione dello stato post-decisione** Si introduce uno stato intermedio osservato dopo la decisione  $x_t$  ma prima della realizzazione di  $\xi_{t+1}$ , detto **stato post-decisione**  $s_t^x$ . Si scinde la transizione in due passi:

$$s_t^x = g_t^{(1)}(s_t, x_t), \quad s_{t+1} = g_{t+1}^{(2)}(s_t^x, \xi_{t+1}). \quad (10-11)$$

**Relazione tra funzioni valore** Si definisce il valore del post-decision state  $V_t^x(s_t^x)$  e vale:

$$V_t^x(s_t^x) = \mathbb{E}[V_{t+1}(s_{t+1}) | s_t^x]. \quad (12)$$

Quindi la ricorsione standard può essere riscritta come ottimizzazione deterministica:

$$V_t(s_t) = \text{opt}_{x_t \in X(s_t)} \{f_t(s_t, x_t) + \gamma V_t^x(s_t^x)\}. \quad (13)$$

**Passo all’indietro e nuovo scambio** Scrivendo la relazione un passo indietro e sostituendo (13), si ottiene una ricorsione che presenta nuovamente lo **scambio** tra **aspettazione** e **ottimizzazione**:

$$V_{t-1}^x(s_{t-1}^x) = \mathbb{E} \left[ \text{opt}_{x_t \in X(s_t)} (f_t(s_t, x_t) + \gamma V_t^x(s_t^x)) \mid s_{t-1}^x \right]. \quad (14)$$

**Connessione con i Q-factors** La coppia **stato–azione**  $(i, a)$  può essere interpretata come **stato post-decisione** prima di osservare la transizione casuale al nuovo stato; la formulazione in termini di **Q-factors** si inserisce naturalmente in questo quadro.

## 10.11 Aumento dello stato nella gestione delle scorte

**Inventario on-hand vs inventario disponibile** L'inventario **on-hand** appare naturale come variabile di stato, ma le decisioni d'ordine dovrebbero basarsi sull'**inventario disponibile**, che include **on-order** e **backlog**.

**Limite dell'equazione di stato standard** L'equazione

$$I_{t+1} = I_t + x_t - d_{t+1}$$

assume disponibilità immediata di quanto ordinato al tempo  $t$ .

**Lead time e variabili di pipeline** Se il **lead time** è un intero  $LT \geq 1$ , si introducono variabili  $z_{t,\tau}$  (pipeline), con  $\tau = 0, 1, \dots, LT - 1$ . Vale:

$$I_{t+1} = I_t + z_{t,0} - d_{t+1}$$

(trascurando l'incertezza) e la decisione  $x_t$  entra come:

$$z_{t+1,LT-1} = x_t.$$

Per  $\tau < LT - 1$ , le transizioni sono uno shift temporale:

$$z_{t+1,\tau} = z_{t,\tau+1}, \quad \tau = 0, 1, \dots, LT - 2. \tag{15}$$

**Items deperibili e stato per età** Per prodotti **deperibili** si introduce un array  $I_{t,\tau}$  che descrive l'inventario per età; le transizioni dipendono dalla politica di issuing **FIFO** o **LIFO**.

## 10.12 Revenue management

**Definizione e obiettivo** Il **revenue management** consiste in modelli e tecniche per massimizzare il **ricavo** ottenuto vendendo **risorse deperibili**; l'idea nasce (come **yield management**) nell'industria aerea.

**Approcci quantity-based e price-based** Esistono due approcci: **quantity-based** (si limita la disponibilità della risorsa per massimizzare il ricavo) e **price-based** (si aggiustano dinamicamente i prezzi, **dynamic pricing**). Si considerano modelli DP quantity-based assumendo costo marginale nullo o trascurabile, così che massimizzare il profitto equivale a massimizzare il ricavo.

**Classi tariffarie e prezzi** Si considerino  $C$  unità di risorsa da allocare a  $n$  classi  $j = 1, \dots, n$ , vendute a prezzi  $p_j$  con:

$$p_1 > p_2 > \dots > p_n.$$

I posti sono fisicamente **identici**, ma differenziati tramite **ancillaries** (cancellazione, vincoli, pasti, ecc.).

**Caratteristiche essenziali** La domanda  $D_j$  è casuale e può realizzarsi con schemi diversi. Due aspetti sono essenziali: **comportamento dei clienti** (segmentazione perfetta vs preferenze da modellare con un **choice model**) e **timing della domanda** (sequentialità auspicabile; caso peggiore quando arrivano prima i clienti low-budget).

**Modelli statici e dinamici; protection levels e bid-prices** Un modello con intervalli disgiunti, ciascuno dedicato a una sola classe, è **statico**; se le richieste sono interleaved nel tempo è **dinamico**. La policy si esprime tramite **protection levels** oppure **bid-prices** (prezzo minimo accettabile per vendere un posto).

### 10.13 Modello statico con segmentazione perfetta della domanda

**Assunzioni, stadi e stato** Le domande sono indipendenti e avvengono sequenzialmente (prima  $D_n$ ); nessun cliente cambia classe. Gli stadi decisionali sono  $j = n, n-1, \dots, 1$ . La variabile di stato è la **capacità residua**  $s_j$ , con stato iniziale  $s_n = C$ . Condizione al contorno:

$$V_0(s_0) = 0, \quad s_0 = 0, 1, \dots, C.$$

**Ricorsione DP in forma scambiata** La ricorsione assume la forma scambiata:

$$V_j(s_j) = \mathbb{E} \left[ \max_{0 \leq x_j \leq \min\{s_j, D_j\}} (p_j x_j + V_{j-1}(s_j - x_j)) \right].$$

Con uno shift di indice e notazione compatta:

$$V_{j+1}(s) = \mathbb{E} \left[ \max_{0 \leq x \leq \min\{s, D_{j+1}\}} (p_{j+1} x + V_j(s - x)) \right]. \quad (16)$$

### 10.14 Valore marginale atteso e livelli di protezione

**Valore marginale atteso della capacità** Si definisce:

$$\Delta V_j(s) := V_j(s) - V_j(s-1),$$

che misura il **costo opportunità** di un posto dato  $s$ .

**Somma telescopica e riscrittura** Si ottiene:

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

La riscrittura usa una **somma telescopica**:

$$V_j(s-x) = V_j(s) - \sum_{z=1}^x \Delta V_j(s+1-z).$$

**Proprietà dei valori marginali** Si possono provare:

$$\Delta V_j(s+1) \leq \Delta V_j(s), \quad \Delta V_{j+1}(s) \geq \Delta V_j(s), \quad \forall s, j.$$

**Regola di accettazione e protezione annidata** I termini  $p_{j+1} - \Delta V_j(s+1-z)$  sono decrescenti in  $z$ : si aumenta  $x$  finché compare il primo termine negativo o si raggiunge  $\min\{s, D_{j+1}\}$ . Si ottiene un livello di protezione **annidato**:

$$y_j^* := \max \{ y : p_{j+1} < \Delta V_j(y) \}, \quad j = 1, \dots, n-1. \quad (18)$$

La decisione ottima allo stadio  $j+1$  è:

$$x_{j+1}^*(s_{j+1}, D_{j+1}) = \min \{ (s_{j+1} - y_j^*)^+, D_{j+1} \}.$$

Lo scambio tra aspettazione e ottimizzazione è giustificato dal fatto che non serve conoscere  $D_{j+1}$  in anticipo: si accetta finché si raggiunge la protezione o si esaurisce la capacità o termina lo stadio.

## 10.15 Modello dinamico con segmentazione perfetta della domanda

**Assunzioni e probabilità di arrivo** Si rilassa la sequenzialità della domanda, mantenendo una segmentazione rigida. Il tempo è discretizzato in  $t = 1, \dots, T$  assumendo **al più un arrivo per intervallo**. Sia  $\lambda_j(t)$  la probabilità di arrivo della classe  $j$  nell'intervallo  $t$ , con vincolo:

$$\sum_{j=1}^n \lambda_j(t) \leq 1, \quad \forall t.$$

**Funzione valore e condizioni al contorno** Si determinano  $V_t(s)$  (capacità residua  $s$ ) con:

$$V_t(0) = 0, \quad t = 1, \dots, T, \quad V_{T+1}(s) = 0, \quad s = 0, 1, \dots, C.$$

**Ricavo disponibile e decisione binaria** Sia  $R(t)$  il **ricavo disponibile** al tempo  $t$ : vale  $p_j$  se arriva un cliente di classe  $j$ , 0 altrimenti. Alla richiesta si decide con  $x \in \{0, 1\}$  se accettare o meno.

**Ricorsione DP in forma scambiata** La ricorsione assume la forma scambiata:

$$V_t(s) = \mathbb{E} \left[ \max_{x \in \{0,1\}} \{R(t)x + V_{t+1}(s-x)\} \right].$$

Usando i **valori marginali attesi** si ottiene una politica in termini di **protection levels**, qui potenzialmente **time-varying**.

## 10.16 Modello dinamico con scelta del cliente

**Modello di scelta** Se la segmentazione perfetta viene rilassata, occorre un **choice model**. La decisione di controllo al tempo  $t$  è il sottoinsieme  $S_t \subseteq N = \{1, \dots, n\}$  delle classi offerte.

**Ricavi e probabilità di scelta** Si introduca  $p_0 = 0$  per il caso di **non-acquisto**. Con probabilità  $\lambda$  arriva un potenziale passeggero; dato  $S_t$ , la probabilità di scelta della classe  $j$  è  $P_j(S_t)$ , includendo  $P_0(S_t)$ , con vincoli:

$$P_j(S) \geq 0, \quad S \subseteq N, \quad j \in S \cup \{0\}, \quad \sum_{j \in S} P_j(S) + P_0(S) = 1, \quad S \subseteq N.$$

**Probabilità di acquisto dipendenti dalla decisione** Le probabilità dipendenti dalla decisione sono  $\lambda P_j(S_t)$  per  $j = 1, \dots, n$ , e  $(1 - \lambda) + \lambda P_0(S_t)$  per  $j = 0$ .

**Ricorsione DP** In questo caso la decisione deve essere dichiarata prima della scelta del passeggero, quindi la ricorsione è nella forma standard:

$$V_t(s) = \max_{S_t \subseteq N} \left\{ \sum_{j \in S_t} \lambda P_j(S_t) (p_j + V_{t+1}(s-1)) + (\lambda P_0(S_t) + 1 - \lambda) V_{t+1}(s) \right\}.$$

**Osservazione conclusiva** La forma della ricorsione riflette la **struttura informativa**: qui la massimizzazione è sull'**aspettazione** perché la decisione precede la scelta del cliente.

# 11 CAP 10 – Programmazione dinamica numerica per stati discreti

## 11.1 Catene di Markov a tempo discreto

**Definizione di catena di Markov a tempo discreto** Una **catena di Markov a tempo discreto** è un processo stocastico con variabile di stato  $s_t$ ,  $t = 0, 1, 2, \dots$ , che assume valori in un insieme **discreto**. Se lo spazio degli stati è numerabile, è possibile associare ciascuno stato a un numero intero e rappresentare il processo mediante una **rete**.

**Probabilità di transizione** Le transizioni tra stati sono rappresentate da archi diretti etichettati con le **probabilità di transizione**. Le probabilità di transizione sono **probabilità condizionate** che dipendono solo dallo stato corrente:

$$\pi(i, j) := \mathbb{P}\{s_{t+1} = j \mid s_t = i\}.$$

**Catene omogenee** Nel caso di problemi a **orizzonte infinito** si utilizzano catene **omogenee** (o **time-invariant**), in cui le probabilità di transizione non dipendono dal tempo.

**Matrice di transizione** Una catena di Markov a tempo discreto può essere descritta raccogliendo le probabilità di transizione in una **matrice di transizione a un passo**  $\Pi$ , in cui l'elemento  $\pi_{ij}$  rappresenta la probabilità di transizione dallo stato  $i$  allo stato  $j$ .

**Proprietà di normalizzazione** Poiché dopo una transizione il processo deve necessariamente trovarsi in uno stato dello spazio degli stati, ogni riga della matrice di transizione somma a uno:

$$\sum_{j=1}^N \pi(i, j) = 1, \quad \forall i.$$

## 11.2 Catene di Markov controllate e MDP

**Azioni e transizioni controllate** Nei **processi decisionali markoviani** (MDP), le transizioni sono parzialmente controllate tramite la selezione di azioni. In ciascuno stato  $i$  è disponibile un insieme finito di azioni ammissibili  $A(i)$ . Per ogni azione  $a \in A(i)$  sono definite probabilità di transizione:

$$\pi(i, a, j).$$

**Distribuzione stazionaria** Per valutare le prestazioni di una politica di controllo stazionario, si può considerare la **probabilità di lungo periodo** di trovarsi in ciascuno stato, indicata con  $q(i)$  e raccolta nel vettore  $q$ . Tali probabilità potrebbero non esistere se la catena non soddisfa opportune proprietà strutturali.

**Proprietà strutturali delle catene** Una catena può presentare stati **transienti**, **ricorrenti**, **assorbenti** e **periodicità**. Si assume una **unichain**, tale che ogni stato possa essere visitato infinitamente spesso nel lungo periodo e raggiunto da ogni altro stato in tempo finito con probabilità positiva. Nel seguito si assume che il processo di Markov sia **ben comportato** per ogni politica di controllo.

## 11.3 MDP a orizzonte temporale finito

**Ricorsione di programmazione dinamica** Per un MDP a **orizzonte finito**, la ricorsione di DP è:

$$V_t(i) = \text{opt}_{a \in A(i)} \left\{ f(i, a) + \sum_j \pi(i, a, j) V_{t+1}(j) \right\}, \quad i \in S.$$

**Cosa insegna l'esempio (arresto ottimo su random walk)** L'esempio mostra che, in un problema di arresto ottimo a orizzonte finito, la politica risultante può essere **non stazionaria** e dipendere dal tempo residuo (confronto tra `wait` e `stop` tramite DP).

## 11.4 Processi decisionali markoviani a orizzonte infinito

**Equazioni di Bellman a orizzonte infinito** Nel caso di un **MDP a orizzonte infinito**, la funzione valore è definita implicitamente da equazioni del tipo:

$$V(i) = \text{opt}_{a \in A(i)} \left\{ f(i, a) + \sum_j \pi(i, a, j) V(j) \right\}, \quad i \in S, \quad (2)$$

oppure, quando il contributo immediato dipende anche dallo stato successivo,

$$V(i) = \text{opt}_{a \in A(i)} \sum_{j \in S} \pi(i, a, j) \{ h(i, a, j) + V(j) \}, \quad i \in S. \quad (3)$$

**Rappresentazione vettoriale** Per MDP finiti, lo spazio degli stati è:

$$S = \{1, \dots, n\},$$

e la funzione valore  $V : S \rightarrow \mathbb{R}$  è rappresentabile come vettore  $V \in \mathbb{R}^n$  con componenti  $V(i)$ .

**Strategie di soluzione** Esistono due strategie fondamentali:

- **Value iteration**: iterazioni economiche, convergenza al valore ottimo solo nel limite;
- **Policy iteration**: iterazioni più costose, ma convergenza in tempo finito per MDP finiti (numero finito di politiche).

## 11.5 Operatori di Bellman

**Operatore ottimo** Data una funzione valore generica  $\tilde{V}$ , si definisce l'operatore  $T$ :

$$[T\tilde{V}](i) = \text{opt}_{a \in A(i)} \sum_{j \in S} \pi(i, a, j) \left\{ h(i, a, j) + \tilde{V}(j) \right\}, \quad i \in S. \quad (4)$$

**Operatore associato a una politica** Data una politica stazionaria generica  $\mu$ , si definisce l'operatore  $T_\mu$ :

$$[T_\mu \tilde{V}](i) = \sum_{j \in S} \pi(i, \mu(i), j) \left\{ h(i, \mu(i), j) + \tilde{V}(j) \right\}, \quad i \in S. \quad (5)$$

**Ruolo degli operatori e punti fissi** L'operatore  $T$  è centrale per la **value iteration**, mentre  $T_\mu$  è centrale per la **policy iteration**. Il vettore valore ottimo  $V$  è un **punto fisso** di  $T$ :

$$V = TV. \quad (6)$$

La funzione valore  $V^\mu$  di una politica stazionaria  $\mu$  è il punto fisso di  $T_\mu$ .

**Condizioni di esistenza** Per un MDP finito con **sconto stretto** ( $\gamma < 1$ ) e contributi per stadio **limitati** (esiste  $M$  con  $|h(i, a, j)| \leq M$ ), si assume che  $T$  e  $T_\mu$  siano **operatori di contrazione**, con punto fisso unico.

**Miglioramento delle politiche** L'operatore  $T$  fornisce una caratterizzazione della politica stazionaria ottima e consente di **migliorare** una politica non ottima usando il valore  $V^\mu$ , che viene ottenuto tramite  $T_\mu$ .

## 11.6 Value iteration

**Idea di base** Con sconto stretto ( $\gamma < 1$ ), si cerca un punto fisso di  $T$  tramite iterazione:

$$V^{(k+1)} = TV^{(k)}.$$

Se  $V^{(k)} \rightarrow V$ , allora  $V$  è un punto fisso di  $T$  e rappresenta la soluzione.

## 11.7 Algoritmo di Value Iteration

---

**Algorithm 1** Value Iteration per MDP a orizzonte infinito

---

- 1: Scegliere una funzione valore iniziale  $V^{(0)}$  (ad esempio  $V^{(0)}(i) = 0$  per ogni  $i \in S$ )
- 2: Scegliere una tolleranza  $\varepsilon$
- 3: Inizializzare  $k = 0$  e **stop=false**
- 4: **while** **stop**  $\neq$  **true** **do**
- 5:     **for** ogni stato  $i \in S$  **do**
- 6:         
$$V^{(k+1)}(i) = \text{opt}_{a \in A(i)} \left\{ f(i, a) + \sum_{j \in S} \pi(i, a, j) V^{(k)}(j) \right\}$$
- 7:     **end for**
- 8:     **if**  $\|V^{(k+1)} - V^{(k)}\|_1 < \varepsilon$  **then**
- 9:         **stop=true**
- 10:     **else**
- 11:          $k = k + 1$
- 12:     **end if**
- 13: **end while**
- 14: Porre  $\hat{V} = V^{(k+1)}$
- 15: Per ogni  $i \in S$ , determinare una politica stimata ottima:

$$\hat{\mu}(i) \in \arg \text{opt}_{a \in A(i)} \left\{ f(i, a) + \sum_j \pi(i, a, j) \hat{V}(j) \right\}$$

- 16: Restituire  $\hat{V}$  e  $\hat{\mu}$
- 

**Cosa insegnà l'esempio numerico (value iteration)** L'esempio illustra che la politica ottima ottenuta da value iteration può dipendere in modo sensibile da probabilità di transizione e fattore di sconto, e che la convergenza in iterazioni può variare significativamente con tali parametri.

## 11.8 Policy iteration

**Motivazione** Con value iteration può accadere di individuare presto la politica ottima (da valori approssimati) ma senza valutarne ancora con precisione il valore. Policy iteration mira invece alla valutazione (più) diretta del valore di una politica.

**Policy evaluation come sistema lineare** Per una politica stazionaria  $\mu$ ,  $T_\mu$  non include ottimizzazione:

$$[T_\mu \tilde{V}](i) = f(i, \mu(i)) + \sum_{j \in S} \pi(i, \mu(i), j) \tilde{V}(j), \quad i \in S.$$

Il valore  $V^\mu$  soddisfa:

$$V^\mu(i) = f(i, \mu(i)) + \sum_{j \in S} \pi(i, \mu(i), j) V^\mu(j), \quad i \in S.$$

Definendo la matrice di transizione indotta  $\Pi^\mu$  e il vettore dei contributi  $f^\mu$ :

$$f^\mu := \begin{bmatrix} f(1, \mu(1)) \\ f(2, \mu(2)) \\ \vdots \\ f(n, \mu(n)) \end{bmatrix}, \quad (9)$$

si ottiene il sistema:

$$V^\mu = f^\mu + \gamma \Pi^\mu V^\mu, \quad (I - \gamma \Pi^\mu) V^\mu = f^\mu.$$

Formalmente:

$$V^\mu = (I - \gamma \Pi^\mu)^{-1} f^\mu.$$

Per matrici grandi e sparse, metodi diretti possono essere proibitivi; si adottano quindi metodi iterativi.

**Policy improvement** Data  $V^\mu$ , una politica migliorata  $\tilde{\mu}$  è definita da:

$$\tilde{\mu}(i) \in \arg \text{opt}_{a \in A(i)} \left\{ f(i, a) + \sum_j \pi(i, a, j) V^\mu(j) \right\}, \quad i \in S.$$

Per un problema di massimizzazione:

$$V^\mu(i) \leq V^{\tilde{\mu}}(i), \quad i \in S,$$

(con disegualanza invertita in minimizzazione). Se  $\mu$  non è ottima, la disegualanza è stretta per almeno uno stato. Poiché il numero di politiche stazionarie deterministiche è finito, una sequenza di miglioramenti porta a una politica ottima.

## 11.9 Algoritmo di Policy Iteration

---

**Algorithm 2** Policy Iteration per MDP finito a orizzonte infinito

---

1: Definire una politica stazionaria iniziale arbitraria  $\mu^{(0)}$

2: Inizializzare  $k = 0$  e `stop=false`

3: **while** `stop`  $\neq$  `true` **do**

4:     **Policy evaluation:** risolvere

$$(I - \gamma \Pi^{\mu^{(k)}}) V^{\mu^{(k)}} = f^{\mu^{(k)}}$$

5:     **Policy improvement:** porre, per ogni  $i \in S$ ,

$$\mu^{(k+1)}(i) \in \arg \text{opt}_{a \in A(i)} \left\{ f(i, a) + \sum_j \pi(i, a, j) V^{\mu^{(k)}}(j) \right\}$$

6:     **if**  $\mu^{(k+1)} = \mu^{(k)}$  **then**

7:         `stop=true`

8:     **else**

9:          $k = k + 1$

10:     **end if**

11: **end while**

12: Restituire la funzione valore ottima e la politica stazionaria ottima

---

**Cosa insegna l'esempio numerico (policy iteration)** L'esempio evidenzia che policy iteration può raggiungere la politica ottima in un numero finito (spesso piccolo) di passi di miglioramento, grazie alla valutazione esplicita della politica corrente.

## 11.10 Value iteration vs. policy iteration

**Confronto e collegamento** Value iteration impiega molte iterazioni economiche e converge nel limite; policy iteration usa poche iterazioni potenzialmente costose e converge in tempo finito per MDP finiti. La valutazione di una politica può anche essere eseguita iterativamente tramite:

$$V_\mu^{(k+1)}(i) = f(i, \mu(i)) + \sum_{j \in S} \pi(i, \mu(i), j) V_\mu^{(k)}(j), \quad i \in S, \quad (10)$$

e interrompendo in modo prematuro tale procedura si ottiene una stima  $\hat{V}_\mu$  da usare nel miglioramento, dando luogo a **optimistic policy iteration** e, più in generale, a metodi di **generalized policy iteration**. Value iteration e policy iteration possono quindi essere visti come estremi di un continuum di approcci.

## 11.11 Collegamento con Reinforcement Learning

**DP model-free e Q-factors** Quando le probabilità di transizione (e possibilmente i contributi immediati) non sono note, si passa a una DP **model-free**. In tale contesto, la funzione valore è tipicamente sostituita dai **Q-factors**  $Q(s, a)$  dipendenti da stati e azioni.

**Q-learning e SARSA** Il **Q-learning** è il corrispettivo RL della value iteration ed è un metodo **off-policy**. I corrispettivi RL della policy iteration sono **on-policy**; un approccio noto è **SARSA**.

**Osservazione operativa in RL** In RL non è in generale possibile valutare esattamente il valore di una politica se ciò richiede simulazioni Monte Carlo costose o esperimenti online; è quindi necessario effettuare un passo di miglioramento prima o poi, generando diverse varianti di strategie.

# 12 CAP 11 – Programmazione dinamica approssimativa e apprendimento per rinforzo per stati discreti

## 12.1 Approximate Dynamic Programming e DP model-free

**Approximate Dynamic Programming** L'Approximate Dynamic Programming (ADP) è il nome collettivo di un'ampia classe di **metodi numerici di programmazione dinamica** che, a differenza dei metodi standard, **non garantiscono** di trovare una **politica ottima**. In cambio della natura approssimata, i metodi ADP mirano ad attenuare le **maledizioni della DP**. Nel seguito si considerano **metodi approssimati** per MDP finiti nel caso a orizzonte infinito.

**Reinforcement Learning** Le versioni **model-free** di **value iteration** e **policy iteration** sono raccolte sotto il termine **Reinforcement Learning (RL)**. Lo schema di base è il seguente: un **agente** interagisce con un sistema per apprendere una **politica di controllo** senza disporre di un **modello** del sistema; al tempo  $t$  osserva lo **stato**  $s_t$ , prova un'**azione**  $a_t$ , e osserva il **contributo immediato**  $h_{t+1}$  e lo **stato successivo**  $s_{t+1}$ . Una buona politica deve bilanciare **obiettivi di breve e lungo periodo**, misurati dal contributo immediato e dal valore dello stato successivo.

## 12.2 Valori di stato, Q-factors e necessità di rappresentazioni compatte

**Perché  $V(s)$  non basta in RL model-free** Un punto chiave nel RL model-free è che, anche con una conoscenza perfetta della **funzione valore  $V(s)$** , non è possibile valutare correttamente la bontà di un'azione se mancano informazioni sulle **transizioni** verso lo stato successivo.

**Q-factors** Un rimedio consiste nell'introdurre i **valori stato–azione**, cioè i **Q-factors**:

$$Q(s, a).$$

Sostituendo i valori di stato con i valori stato–azione, diventa possibile adottare **strategie di campionamento** per apprendere una politica decisionale. Si assume che lo spazio **stato–azione** non sia troppo grande e che una rappresentazione **tabellare** dei Q-factors sia fattibile; spesso, però, ciò non vale, e risulta necessaria una **rappresentazione compatta**.

## 12.3 Campionamento e stima in contesti non stazionari

**Monte Carlo** In teoria della probabilità si è interessati al valore atteso di una funzione  $h(\cdot)$  di una variabile aleatoria  $X$ :

$$\theta = \mathbb{E}[h(X)] = \int_{\mathcal{X}} h(x) f_X(x) dx.$$

Poiché l'integrazione ad alta dimensionalità è difficile, si ricorre al **campionamento Monte Carlo**:

$$\hat{\theta} = \frac{1}{m} \sum_{k=1}^m h(X^{(k)}) ,$$

che fornisce una stima **non distorta** di  $\theta$ .

**Complicazioni negli MDP** Negli MDP si vogliono stimare **valori di stato  $V(i)$**  o **Q-factors  $Q(i, a)$** . In questo contesto l'apprendimento statistico deve affrontare:

- il bilanciamento tra **esplorazione** ed **exploitation**;
- la difficoltà di stimare un **target non stazionario**.

## 12.4 Il tradeoff esplorazione–exploitation

**Scelta dell'azione durante l'apprendimento** Se si sta imparando  $Q(i, a)$  e ci si trova nello stato  $i$ , una strategia di pura **exploitation** sceglierrebbe l'azione che massimizza le stime correnti. Questa scelta è sensata solo quando le stime dei valori delle azioni sono sufficientemente accurate, condizione che tipicamente non vale all'inizio. Un'azione può apparire non buona perché:

- il contributo immediato è stato stimato male;
- conduce a uno stato molto buono che non è ancora stato esplorato.

Da qui nasce il problema fondamentale di bilanciare **exploitation** ed **esplorazione**.

**Problema statico semplificato** Si consideri un insieme finito  $A$  di azioni, ciascuna associata a una ricompensa casuale  $R(a)$ . Se fossero noti i valori attesi  $\nu(a)$ , un decisore neutrale al rischio sceglierrebbe l'azione con valore atteso massimo. Se invece si hanno solo stime rumorose  $\hat{\nu}(a)$ , la pura exploitation (massimizzare  $\hat{\nu}$ ) può essere una strategia inadeguata.

## 12.5 Multi-armed bandits

**Idea** Un classico esempio è il **multi-armed bandit**: una slot machine con bracci diversi, associati a differenti distribuzioni di payoff. Con pochi campioni è facile mancare l'azione ottima, soprattutto quando l'azione con valore atteso più alto presenta anche una varianza elevata.

## 12.6 Politiche randomizzate per l'esplorazione

**Compromesso tra estremi** All'estremo opposto della pura exploitation, la pura esplorazione seleziona azioni a caso con probabilità uniformi. Occorre un compromesso, che porta a **politiche randomizzate**.

**Approccio  $\varepsilon$ -greedy** **Versione statica:** fissata una probabilità  $\varepsilon$ , con probabilità  $1 - \varepsilon$  si seleziona l'azione più promettente, e con probabilità  $\varepsilon$  si seleziona un'azione casuale.

**Versione dinamica:** una possibile raffinazione è rendere  $\varepsilon$  dipendente dal tempo, ad esempio:

$$\varepsilon(k) = \frac{c}{d+k}, \quad \text{oppure} \quad \varepsilon(k) = d + \frac{c}{k}.$$

**Esplorazione di Boltzmann** L'esplorazione può essere realizzata tramite una funzione soft-max che definisce la probabilità di scegliere l'azione  $a \in A$ :

$$\pi(a) = \frac{\exp(\rho \hat{\nu}(a))}{\sum_{a' \in A} \exp(\rho \hat{\nu}(a'))}.$$

Quando  $\rho = 0$  si ha **pura esplorazione**, mentre per  $\rho \rightarrow +\infty$  si tende alla **pura exploitation**.

## 12.7 Non stazionarietà e smoothing esponenziale

**Media campionaria e aggiornamento incrementale** Per stimare  $\theta = \mathbb{E}[X]$  si usa la media campionaria, che soddisfa:

$$\hat{\theta}(m) = \hat{\theta}(m-1) + \frac{1}{m} (X(m) - \hat{\theta}(m-1)).$$

La correzione è smussata da  $1/m$ , che tende a zero all'aumentare di  $m$ .

**Passo costante in contesti non stazionari** In un contesto non stazionario si può mantenere costante la correzione:

$$\boxed{\hat{\theta}(m) = \hat{\theta}(m-1) + \alpha (X(m) - \hat{\theta}(m-1)) = \alpha X(m) + (1-\alpha)\hat{\theta}(m-1)}$$

dove  $\alpha \in (0, 1)$  pesa nuova informazione e informazione passata.

**Smoothing esponenziale e ruolo di  $\alpha$**  Questa procedura è nota come **smoothing esponenziale**: le osservazioni più vecchie ricevono pesi **esponenzialmente decrescenti**. Il coefficiente  $\alpha$  è detto **learning rate**, **smoothing coefficient** o **forgetting factor**. Un  $\alpha$  grande produce aggiornamenti più reattivi e “nervosi”; un  $\alpha$  piccolo introduce più inerzia e filtraggio del rumore.

**Learning rate decrescente** Se la non stazionarietà è dovuta solo all'apprendimento (e non a variazioni della dinamica del sistema), si può usare una sequenza decrescente  $\alpha(k)$  che soddisfa condizioni del tipo:

$$\boxed{\sum_{k=1}^{\infty} \alpha(k) = \infty, \quad \sum_{k=1}^{\infty} \alpha(k)^2 < \infty.}$$

## 12.8 Apprendimento tramite differenze temporali e SARSA

**Operatore  $T_\mu$  e punto fisso** Si introduce l'operatore  $T_\mu$ :

$$[T_\mu \tilde{V}](i) = \sum_{j \in S} \pi(i; \mu(i); j) \left( h(i; \mu(i); j) + \gamma \tilde{V}(j) \right), \quad i \in S,$$

dove  $\mu$  è una **politica stazionaria** e  $\tilde{V}$  è una funzione valore. Il valore  $V_\mu$  della politica  $\mu$  si ottiene trovando il **punto fisso** di  $T_\mu$ :

$$T_\mu V_\mu = V_\mu.$$

**Punto fisso in termini di Q-factors della politica** La stessa idea può essere riscritta in termini dei **Q-factors** associati alla politica stazionaria  $\mu$ :

$$Q_\mu(i; \mu(i)) = \mathbb{E} \left[ h(i; \mu(i); j) + \gamma Q_\mu(j; \mu(j)) \right],$$

dove l'aspettativa è rispetto allo **stato successivo**  $j$ .

## 12.9 Schema di punto fisso, smoothing e differenze temporali

**Riformulazione astratta** Si consideri un'equazione di punto fisso:

$$y = Hy,$$

dove  $y$  è un vettore e  $H$  è un operatore. L'iterazione diretta  $y^{(k)} = Hy^{(k-1)}$  non garantisce convergenza in generale e, inoltre, in pratica  $H$  non è valutabile esattamente perché richiede aspettative non note.

**Schema incrementale con learning rate** L'equazione può essere riscritta come:

$$y = y + \alpha(Hy - y),$$

da cui lo schema iterativo:

$$y^{(k)} = y^{(k-1)} + \alpha(Hy^{(k-1)} - y^{(k-1)}).$$

**Sostituzione dell'aspettativa con osservazioni** Applicando **smoothing esponenziale** e sostituendo l'aspettativa con osservazioni campionate, si ottiene lo schema tipico dell'apprendimento statistico.

**Differenza temporale e aggiornamento (SARSA)** Quando si osserva una transizione da  $i = s(k)$  a  $j = s(k+1)$  applicando l'azione prescritta dalla politica  $\mu$ , la **differenza temporale** è:

$$\delta(k) = \left[ h(i; \mu(i); j) + \gamma \hat{Q}_\mu^{(k-1)}(j; \mu(j)) \right] - \hat{Q}_\mu^{(k-1)}(i; \mu(i))$$

e l'aggiornamento dei Q-factors è:

$$\hat{Q}_\mu^{(k)}(i; \mu(i)) = \hat{Q}_\mu^{(k-1)}(i; \mu(i)) + \alpha \delta(k).$$

La differenza temporale svolge il ruolo della correzione  $Hy - y$ . Poiché si usano stime per ottenere nuove stime, il termine **bootstrapping** descrive questo tipo di procedura.

**Significato di SARSA e natura on-policy** L'algoritmo è noto come **SARSA** perché utilizza la sequenza (**State, Action, Reward, State, Action**): si osserva lo stato  $i$ , si seleziona l'azione  $\mu(i)$ , si osservano il reward e il nuovo stato  $j$ , e si considera l'azione  $\mu(j)$  nel nuovo stato. SARSA è un approccio **on-policy**: si agisce secondo una politica e si apprende il valore di quella stessa politica, come richiesto nella **policy iteration**.

**Esplorazione e policy iteration generalizzata (ottimistica)** Servono due ingredienti:

- una quantità adeguata di **esplorazione**, ad esempio tramite  **$\varepsilon$ -greedy** (con probabilità  $\varepsilon(k)$  si devia dalla politica e si seleziona un'azione casuale);
- l'interruzione dell'apprendimento prima della valutazione esatta della politica, seguita da un passo di **miglioramento** della politica.

Questo schema è detto **generalized o optimistic policy iteration**: si assume ottimisticamente che la politica corrente sia stata valutata con accuratezza sufficiente. Ciò può accelerare l'apprendimento, ma può essere dannoso per la convergenza.

**Miglioramento greedy della politica** Quando si dispone di un'approssimazione  $\hat{Q}_\mu(i; a)$ , si tenta di migliorare la politica con una scelta greedy:

$$\tilde{\mu}(i) \in \arg \underset{a \in A(i)}{\text{opt}} \hat{Q}_\mu(i; a), \quad i \in S.$$

A differenza dell'iterazione esatta, non vi è garanzia che  $\tilde{\mu}$  sia effettivamente migliorativa.

## 12.10 Q-learning per MDP finiti

**Equazioni per i Q-factors ottimali** Nel Q-learning si considera direttamente la forma DP basata sui **Q-factors ottimali**:

$$Q(i; a) = \sum_{j \in S} \pi(i; a; j) \left[ h(i; a; j) + \gamma \underset{a' \in A(j)}{\text{opt}} Q(j; a') \right], \quad i \in S, a \in A(i),$$

$$V(i) = \underset{a \in A(i)}{\text{opt}} Q(i; a).$$

A differenza di SARSA, qui non si valuta una politica fissata  $\mu$ , ma si mira ai fattori ottimali.

**Logica off-policy e scelta dell'azione** In Q-learning la politica è **implicita** nei Q-factors: aggiornandoli continuamente, si continua a cambiare politica; in un certo senso, si usa una politica per impararne un'altra. Al passo  $k$ , nello stato  $s(k) = i$ , la scelta greedy è:

$$a(k) \in \arg \underset{a \in A(i)}{\text{opt}} \hat{Q}^{(k-1)}(i; a).$$

**Target campionario e aggiornamento con smoothing** Dopo l'azione si osservano lo stato successivo  $j = s(k+1)$  e il contributo immediato. Si costruisce un target campionario che bilancia breve e lungo periodo:

$$\tilde{q} = h(i; a(k); j) + \gamma \underset{a' \in A(j)}{\text{opt}} \hat{Q}^{(k-1)}(j; a').$$

L'aggiornamento tramite smoothing esponenziale è:

$$\hat{Q}^{(k)}(s(k); a(k)) = \alpha \tilde{q} + (1 - \alpha) \hat{Q}^{(k-1)}(s(k); a(k)).$$

**Forma a differenze temporali** La correzione può essere scritta come differenza temporale:

$$\delta(k) = \left[ h(s(k); a(k); j) + \gamma \underset{a' \in A(j)}{\text{opt}} \hat{Q}^{(k-1)}(j; a') \right] - \hat{Q}^{(k-1)}(s(k); a(k)),$$

e l'aggiornamento riguarda solo la coppia stato–azione osservata.

**Confronto SARSA vs Q-learning** La differenza si coglie confrontando le correzioni:

- **SARSA** è **on-policy** e riferito a una politica stazionaria incumbente  $\mu$ ;
- **Q-learning** è **off-policy** e la politica cambia mentre si aggiornano i Q-factors.

## 13 CAP 12 – Simulazione

Scrivere un paragrafo per fare recap dei problemi affrontati di simulazione.

- Cap 1 pag 4/48 simulazione M/M/1