

SAPIENZA
UNIVERSITÀ DI ROMA

Modellazione PL

Modelli di programmazione lineare

Dispensa didattica con modelli, esempi numerici svolti, casi di studio, implementazioni complete in Python/Gurobi e analisi di sensitività. Modelli di programmazione lineare: produzione, ricarica di veicoli elettrici, DEA; decisioni sotto incertezza (Newsvendor, VaR e CVaR, arbitraggio); regressione robusta e quantile.

Fabio Furini
fabiofurini.github.io
Dipartimento di Ingegneria informatica, automatica e gestionale — Sapienza Università di Roma

Versione online e materiale scaricabile:
fabiofurini.github.io/modellazione-pl/
github.com/fabiofurini/modellazione-pl

Anno accademico 2026–2027

Indice

I Strumenti **4**

1 Introduzione alla modellazione PL **5**
1.1 Che cosa faremo . . . 5
1.2 Modelli lineari e sigle usate . . . 6
1.3 Notazione . . . 6
1.4 Una scelta di campo: solo variabili continue . . . 6
1.5 Mappa dei modelli . . . 7
1.6 Struttura di ogni capitolo . . . 7
1.7 Organizzazione del corso . . . 7
1.8 In sintesi . . . 8

2 Richiami: LP e analisi di sensitività **9**
2.1 Teoria della programmazione lineare . . . 9
2.1.1 La dualità . . . 9
2.1.2 Le condizioni di ottimalità: gli scarti complementari . . . 12
2.1.3 L'analisi di sensitività e i prezzi ombra . . . 15
2.2 Il protocollo di analisi di sensitività . . . 17
2.3 In sintesi . . . 17

3 Il solver: costruire, risolvere, interpretare **18**
3.1 I comandi generali . . . 18
3.1.1 Installazione e licenza . . . 18
3.1.2 I cinque passi per costruire un modello . . . 19
3.1.3 Far girare il modello e leggere il log . . . 19
3.1.4 Recuperare la soluzione . . . 20
3.2 I comandi per i modelli lineari . . . 20
3.2.1 Prezzi ombra, costi ridotti e intervalli: gli attributi . . . 20
3.2.2 Interpretare l'output: la checklist . . . 20
3.2.3 Due letture complete . . . 21
3.3 In sintesi . . . 23

II Modelli deterministici **24**

4 Produzione e scorte multiperiodali **25**
4.1 Motivazione gestionale . . . 25
4.2 Costruzione guidata del modello . . . 26

4.3 Esempio svolto a mano . . . 27
4.4 Caso di studio . . . 27
4.5 Implementazione . . . 28
4.6 Risultati . . . 28
4.7 Analisi di sensitività . . . 29
4.8 In sintesi . . . 31
4.9 Esercizi . . . 31

5 Ricarica intelligente di veicoli elettrici 32
5.1 Motivazione gestionale . . . 32
5.2 Costruzione guidata del modello . . . 33
5.3 Esempio svolto a mano . . . 34
5.4 Caso di studio . . . 34
5.5 Risultati . . . 35
5.6 Analisi di sensitività . . . 36
5.7 In sintesi . . . 37
5.8 Esercizi . . . 37

6 Data Envelopment Analysis 38
6.1 Motivazione gestionale . . . 38
6.2 Unità, input, output ed efficienza . . . 39
6.3 Il modello CCR orientato all'input . . . 39
6.4 La forma di inviluppo: il duale . . . 41
6.5 Il modello BCC e l'efficienza di scala . . . 42
6.6 Scarti ed efficienza debole e forte . . . 43
6.7 Esempio svolto a mano . . . 44
6.8 Caso di studio . . . 45
6.9 Implementazione . . . 45
6.10 Risultati . . . 46
6.11 Analisi di sensitività . . . 48
6.12 In sintesi . . . 49
6.13 Esercizi . . . 49

III Decisioni sotto incertezza 51

7 Il Newsvendor e le sue varianti 52
7.1 Motivazione gestionale . . . 52
7.2 Il modello base . . . 53
7.3 Esempio svolto a mano (domanda discreta) . . . 54
7.4 Caso di studio: la regola del quantile in continuo . . . 54
7.5 La formulazione lineare a scenari . . . 55
7.6 Livelli di servizio . . . 56
7.7 Avversione al rischio: media-CVaR . . . 56
7.8 Multiprodotto con budget condiviso . . . 57
7.9 In sintesi . . . 57
7.10 Esercizi . . . 58

8 VaR e CVaR **59**
8.1 Motivazione gestionale . . . 59
8.2 Definizioni e proprietà . . . 60
8.3 Esempio svolto a mano . . . 61
8.4 Caso di studio 1: portafoglio mean-CVaR . . . 61
8.5 Caso di studio 2: supply chain a due stadi . . . 62
8.6 Analisi di sensitività . . . 63
8.7 In sintesi . . . 64
8.8 Esercizi . . . 64

9 Arbitraggio e prezzatura **66**
9.1 Motivazione gestionale . . . 66
9.2 Costruzione guidata del modello . . . 67
9.3 La dualità è la prezzatura: le probabilità neutrali al rischio . . . 68
9.4 Esempio svolto a mano . . . 69
9.5 Caso di studio . . . 69
9.6 Risultati . . . 70
9.7 Analisi di sensitività: l'intervallo dei prezzi coerenti . . . 70
9.8 In sintesi . . . 71
9.9 Esercizi . . . 71

IV Ottimizzazione e machine learning **72**

10 Regressione robusta e quantile **73**
10.1 Motivazione gestionale . . . 73
10.2 Il modello . . . 74
10.3 Caso di studio: lo storico della panetteria . . . 75
10.4 La lettura duale: i punti di appoggio . . . 76
10.5 Regressione quantile e scorta di sicurezza . . . 78
10.6 Selezione delle caratteristiche con un budget . . . 79
10.7 In sintesi . . . 80
10.8 Esercizi . . . 81

V Organizzazione del laboratorio **82**

11 Organizzazione del laboratorio **83**
11.1 Percorso essenziale in quattro laboratori . . . 83
11.2 Errori ricorrenti da evitare . . . 83
11.3 Materiali del corso . . . 84

Parte I

Strumenti

Capitolo 1

Introduzione alla modellazione PL

Prerequisiti

- Dal modulo *Programmazione lineare*, Capitolo 4 «Problemi di programmazione lineare»: la forma di un problema di programmazione lineare: funzione obiettivo, vincoli lineari, domini delle variabili.
- Dal modulo *Programmazione lineare*, Capitolo 16 «Tecniche di modellazione per la programmazione lineare»: le tecniche di base per tradurre un enunciato in un modello di programmazione lineare.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- descrivere il percorso di ogni capitolo: problema gestionale, modello, Python e solver, interpretazione;
- usare la notazione del corso per dati, variabili, duali, costi ridotti e scarti;
- spiegare perché il corso usa solo variabili continue e che cosa se ne guadagna;
- orientarsi nella mappa dei modelli e nella struttura di ogni capitolo.

1.1 Che cosa faremo

Questo corso insegna a usare la programmazione lineare come strumento di *decisione*, non solo di calcolo. Ogni capitolo parte da un problema gestionale concreto — quanto produrre, come ricaricare una flotta, quanto è efficiente una filiale, quanto rischio accettare — e lo trasforma in un modello di programmazione lineare che viene poi implementato in Python, risolto con un solver e, soprattutto, *interrogato*: quanto vale un'ora di capacità in più? La soluzione resiste se i dati cambiano del 5%? Quale compromesso consiglieremmo a un decisore?

> **La domanda giusta**
>
> Alla fine di ogni esercitazione la domanda non è soltanto *"qual è l'ottimo?"*, ma *"quale decisione suggeriamo e quanto è robusta?"*. Il valore di un modello sta nella qualità della raccomandazione che permette di formulare.

1.2 Modelli lineari e sigle usate

Tutti i modelli della dispensa sono **LP** (*Linear Programming*, programmazione lineare): obiettivo e vincoli sono funzioni lineari delle variabili. Un LP è sempre un problema convesso: ogni ottimo locale è anche globale, e la dualità del capitolo 2 vale senza eccezioni. Le altre sigle ricorrenti sono definite al primo uso nel rispettivo capitolo.

Prerequisito del corso è il modulo *Programmazione lineare*, in particolare la dualità e l'analisi di sensitività. I modelli con obiettivo quadratico o non lineare sono nel modulo *Modellazione PNL*.

1.3 Notazione

In tutta la dispensa vale la stessa convenzione:

- **scalari e indici**: lettere minuscole in corsivo (x_{it}, b_t, λ, i, t);
- **insiemi di oggetti indicizzati dai numeri**: gli oggetti di un modello (prodotti, canali, titoli, veicoli, quartieri, scenari) sono numerati e gli indici corrono su insiemi enumerati esplicitamente, $i \in \{1, 2, \ldots, n\}$ — così la notazione del testo è già quella del modello costruito; i *conteggi* sono interi positivi ($n \in \mathbb{Z}_{\geq 1}$);
- **dati numerici**: valori razionali, $\mathbb{Q}_{>0}$ o $\mathbb{Q}_{\geq 0}$ (i dati di un'istanza reale sono sempre numeri di macchina);
- **vettori**: lettere minuscole in **grassetto** ($\boldsymbol{x}$, $\boldsymbol{w}$); tutti i vettori sono colonna e $\boldsymbol{a}'\boldsymbol{b}$ indica il prodotto scalare;
- **matrici**: lettere maiuscole in **grassetto** ($\boldsymbol{A}$);
- **grafi**: $G = (N, A)$ con nodi N e archi A (gli unici insiemi non enumerati, per loro natura);
- **variabili duali**: π_i, una per vincolo; **costi ridotti**: $\bar{c}_j$; **scarti**: $\bar{s}_i$;
- **soluzioni**: le variabili incognite si scrivono lisce (x_j, π_i); la *barra* indica i valori di una soluzione *ammissibile* ($\bar{x}_j$, $\bar{\pi}_i$, $\bar{s}_i$, $\bar{c}_j$), la *tilde* quelli di una soluzione *ottima* ($\tilde{x}_j$, $\tilde{\pi}_i$, $\tilde{z}$).

Ogni coefficiente e ogni insieme è *definito prima di essere usato*. In ogni modello: (i) le variabili sono introdotte esplicitamente prima della formulazione; (ii) obiettivo e ciascuna famiglia di vincoli hanno un numero d'equazione; (iii) i vincoli di non negatività, che definiscono le variabili, chiudono il modello; (iv) dopo il modello, un elenco descrive obiettivo e vincoli famiglia per famiglia, con il numero di vincoli di ciascuna.

1.4 Una scelta di campo: solo variabili continue

Tutti i modelli di questa dispensa usano **esclusivamente variabili continue**: niente variabili binarie o intere. È una scelta deliberata, con tre motivazioni.

1. **Didattica**: con variabili continue valgono la dualità, i prezzi ombra, i costi ridotti — gli strumenti che trasformano una soluzione numerica in una spiegazione economica. Con variabili binarie questi strumenti si perdono o si indeboliscono.
2. **Copertura**: una quantità sorprendente di decisioni reali è naturalmente continua: quantità, flussi, quote di portafoglio, potenze, pesi.

3. **Progressione**: gli stessi strumenti — un LP, il suo duale, i prezzi ombra — servono per la produzione, la finanza, la statistica, senza cambiare solver.

1.5 Mappa dei modelli

Cap.	Applicazione	Classe	Concetto centrale
4	Produzione e scorte multiperiodali	LP	dualità, capacità
5	Ricarica di veicoli elettrici	LP	energia, picchi
6	Data Envelopment Analysis	LP	efficienza relativa
7	Newsvendor e varianti	LP stoc.	decidere prima di sapere
8	VaR e CVaR	LP	rischio di coda
9	Arbitraggio e prezzatura	LP	la dualità che prezza
10	Regressione robusta e quantile	LP	stimare con un LP

1.6 Struttura di ogni capitolo

Ogni capitolo applicativo segue la stessa scaletta in otto passi, pensata per essere replicata nelle relazioni di laboratorio:

1. **Motivazione gestionale** — il problema raccontato come decisione aziendale;
2. **Costruzione guidata del modello** — insiemi, parametri, variabili, vincoli e obiettivo, spiegati riga per riga;
3. **Esempio numerico svolto a mano** — un'istanza minima risolta con tutti i passaggi, per capire il meccanismo prima di accendere il solver;
4. **Caso di studio** — un'istanza realistica con dati riproducibili (CSV);
5. **Implementazione Python/Gurobi** — il codice completo, commentato;
6. **Risultati** — l'output del solver tradotto in linguaggio manageriale;
7. **Analisi di sensitività** — prezzi ombra, scenari, trade-off;
8. **Esercizi** — dalla verifica alla piccola estensione di ricerca.

> I numeri riportati in ogni capitolo sono generati dagli script della cartella `python/`: eseguendo `python3 esegui_tutti.py` si rigenerano dati, tabelle e figure. Se modificate i dati, i numeri della dispensa vanno rigenerati.

1.7 Organizzazione del corso

Il percorso essenziale prevede quattro laboratori (capitolo 11 per il dettaglio): (1) produzione e scorte con dualità e prezzi ombra; (2) Data Envelopment Analysis e forma di inviluppo; (3) decisioni a scenari (Newsvendor o CVaR); (4) progetto a scelta con presentazione manageriale.

1.8 In sintesi

In sintesi

- Ogni capitolo va da un problema gestionale a un modello, a Python e al solver; il modello non restituisce solo un numero, va interrogato.
- La barra indica i valori di una soluzione ammissibile, la tilde quelli di una ottima; π_i sono i duali, $\bar{c}_j$ i costi ridotti, $\bar{s}_i$ gli scarti.
- Solo variabili continue, per scelta: dualità, prezzi ombra e costi ridotti valgono allora in ogni modello.
- Ogni capitolo segue lo stesso schema: motivazione, modello, esempio a mano, caso di studio, codice, risultati, sensitività ed esercizi.

Capitolo 2

Richiami: programmazione lineare e analisi di sensitività

Prerequisiti

- Dal modulo *Programmazione lineare*, Capitolo 9 «Teoria della dualità»: il duale di un problema di programmazione lineare, la dualità debole e forte, gli scarti complementari.
- Dal modulo *Programmazione lineare*, Capitolo 14 «Analisi di sensitività»: prezzi ombra, costi ridotti e i loro intervalli di validità.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- scrivere il duale di un LP in forma generale con le regole di conversione;
- verificare l'ottimalità con gli scarti complementari;
- dedurre il segno di un prezzo ombra dal verso del vincolo e dal verso dell'obiettivo;
- leggere un prezzo ombra e un costo ridotto insieme ai loro intervalli di validità;
- applicare i sei passi del protocollo di sensitività.

Questo capitolo raccoglie gli strumenti teorici della programmazione lineare usati in tutta la dispensa: dualità, condizioni di ottimalità e analisi di sensitività. Chiude il protocollo di sensitività usato in ogni laboratorio. È un riferimento rapido: la teoria completa, con le dimostrazioni e gli algoritmi, è nel modulo *Programmazione lineare*, ai capitoli «Teoria della dualità» e «Analisi di sensitività».

2.1 Teoria della programmazione lineare

2.1.1 La dualità

Chiamiamo $M = \{1, 2, \dots, m\}$ l'insieme degli indici dei vincoli e $N = \{1, 2, \dots, n\}$ quello degli indici delle variabili. I *dati* del problema sono i costi $c_j \in \mathbb{R}$ per $j \in N$, i coefficienti $a_{ij} \in \mathbb{R}$ e i termini noti $b_i \in \mathbb{R}$ per $i \in M$. Gli insiemi $M_{\leq}$, $M_{=}$ e $M_{\geq}$ ripartiscono M secondo il verso del vincolo; gli insiemi $N_{\geq 0}$, $N_{\gtrless 0}$ e $N_{\leq 0}$ ripartiscono N secondo il segno della variabile (non

negative, libere, non positive). Le *variabili* sono le x_j del primale e, nel duale, una π_i per ciascuno degli m vincoli del primale.

Il primale (P), forma generale

$$
\begin{aligned}
\text{(P)} \qquad \max \quad & \sum_{j \in N} c_j \, x_j & & \\
\text{soggetto a} \quad & \sum_{j \in N} a_{ij} \, x_j \leq b_i, & & \forall i \in M_{\leq}, \\
& \sum_{j \in N} a_{ij} \, x_j = b_i, & & \forall i \in M_{=}, \\
& \sum_{j \in N} a_{ij} \, x_j \geq b_i, & & \forall i \in M_{\geq}, \\
& x_j \geq 0, & & \forall j \in N_{\geq 0}, \\
& x_j \gtreqless 0, & & \forall j \in N_{\gtreqless 0}, \\
& x_j \leq 0, & & \forall j \in N_{\leq 0}.
\end{aligned}
$$

Il duale (D), forma generale

$$
\begin{aligned}
\text{(D)} \qquad \min \quad & \sum_{i \in M} b_i \, \pi_i & & \\
\text{soggetto a} \quad & \sum_{i \in M} a_{ij} \, \pi_i \geq c_j, & & \forall j \in N_{\geq 0}, \\
& \sum_{i \in M} a_{ij} \, \pi_i = c_j, & & \forall j \in N_{\gtreqless 0}, \\
& \sum_{i \in M} a_{ij} \, \pi_i \leq c_j, & & \forall j \in N_{\leq 0}, \\
& \pi_i \geq 0, & & \forall i \in M_{\leq}, \\
& \pi_i \gtreqless 0, & & \forall i \in M_{=}, \\
& \pi_i \leq 0, & & \forall i \in M_{\geq}.
\end{aligned}
$$

Le regole della dualità

Ogni vincolo del primale genera una variabile duale, ogni variabile del primale un vincolo duale, con le corrispondenze:

Primale (max)	Duale (min)
vincolo $\leq b_i \quad (i \in M_{\leq})$	variabile $\pi_i \geq 0$
vincolo $= b_i \quad (i \in M_{=})$	variabile $\pi_i \gtreqless 0$
vincolo $\geq b_i \quad (i \in M_{\geq})$	variabile $\pi_i \leq 0$
variabile $x_j \geq 0 \quad (j \in N_{\geq 0})$	vincolo $\geq c_j$
variabile $x_j \gtreqless 0 \quad (j \in N_{\gtreqless 0})$	vincolo $= c_j$
variabile $x_j \leq 0 \quad (j \in N_{\leq 0})$	vincolo $\leq c_j$

Due letture veloci: le *uguaglianze* del primale danno variabili duali *libere* (e viceversa); in un massimo un vincolo $\leq$ di risorsa ha $\pi_i \geq 0$ (più risorsa, più valore). Per un primale di *minimo* la tabella si legge da destra a sinistra.

La coppia è legata dai due teoremi della dualità.

Teorema della dualità debole

Per ogni soluzione ammissibile $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ del primale (P) e ogni soluzione ammissibile $(\bar{\pi}_1, \bar{\pi}_2, \ldots, \bar{\pi}_m)$ del duale (D) vale

$$\sum_{j \in N} c_j \, \bar{x}_j \;\leq\; \sum_{i \in M} b_i \, \bar{\pi}_i.$$

Teorema della dualità forte

Il primale (P) ha una soluzione ottima $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ *se e solo se* il duale (D) ha una soluzione ottima $(\tilde{\pi}_1, \tilde{\pi}_2, \ldots, \tilde{\pi}_m)$, e in tal caso i due valori ottimi coincidono:

$$\sum_{j \in N} c_j \, \tilde{x}_j \;=\; \sum_{i \in M} b_i \, \tilde{\pi}_i.$$

Spiegazione riga per riga

La dualità debole dice che ogni soluzione ammissibile del duale è un *limite superiore* per il valore del primale di massimo (e ogni soluzione ammissibile del primale un limite inferiore per il duale): se due soluzioni ammissibili hanno lo stesso valore, sono entrambe ottime. La dualità forte garantisce che all'ottimo questo accade sempre: il divario si chiude. Esistenza non significa unicità: gli ottimi possono essere più d'uno, ma i valori ottimi coincidono. Le variabili duali π_i e i costi ridotti che presto definiremo sono i numeri che useremo di continuo.

Costruire il duale con le regole: l'LP 2×2

Applichiamo le regole all'LP 2×2, che è proprio nella forma canonica di massimo (vincoli $\leq$, variabili non negative):

$$\begin{aligned} \max \quad & 30\,x_1 + 50\,x_2 & & \\ \text{soggetto a} \quad & x_1 + 3\,x_2 & \leq & \; 90 \\ & 2\,x_1 + x_2 & \leq & \; 80 \\ & x_1, \quad x_2 & \geq & \; 0 \end{aligned}$$

- **Passo 1 — ogni vincolo genera una variabile duale.** Due vincoli $\leq$ in un massimo $\Rightarrow \pi_1, \pi_2 \geq 0$.
- **Passo 2 — ogni variabile genera un vincolo duale.** I coefficienti sono le *colonne* del primale, i termini noti i costi: da $x_1 \geq 0$ nasce $\pi_1 + 2\pi_2 \geq 30$; da $x_2 \geq 0$ nasce $3\pi_1 + \pi_2 \geq 50$.
- **Passo 3 — l'obiettivo scambia i ruoli**: i termini noti $(90, 80)$ diventano i costi del duale, che è un problema di *minimo*.

$$\begin{array}{rrrrrrr} \min & 90\,\pi_1 & + & 80\,\pi_2 & & \\ \text{soggetto a} & \pi_1 & + & 2\,\pi_2 & \geq & 30 \\ & 3\,\pi_1 & + & \pi_2 & \geq & 50 \\ & \pi_1, & & \pi_2 & \geq & 0 \end{array}$$

Costruire il duale con le regole: tutti i casi in un solo LP

Costruiamo ora un LP di *minimo* che contiene tutti i casi della tabella — tre variabili con i tre segni e tre vincoli generici con i tre versi; essendo un minimo, la tabella si legge da destra a sinistra:

$$\begin{array}{rrrrrrrr} \min & 5\,x_1 & + & 8\,x_2 & - & 9\,x_3 & & \\ \text{soggetto a} & x_1 & + & x_2 & & & \geq & 30 \\ & x_1 & + & x_2 & - & x_3 & = & 100 \\ & x_1 & - & 2\,x_2 & & & \leq & -20 \\ & x_1 & & & & & \geq & 0 \\ & & & x_2 & & & \gtreqless & 0, \\ & & & & & x_3 & \leq & 0 \end{array}$$

- **Passo 1 — le variabili duali**, una per vincolo, con il segno dato dal verso: $\pi_1 \geq 0$ (vincolo $\geq$ in un minimo), $\pi_2 \gtreqless 0$ (uguaglianza), $\pi_3 \leq 0$ (vincolo $\leq$ in un minimo).
- **Passo 2 — i vincoli duali**, uno per variabile, con il verso dato dal segno: da $x_1 \geq 0$ nasce $\pi_1 + \pi_2 + \pi_3 \leq 5$; da $x_2 \gtreqless 0$ nasce $\pi_1 + \pi_2 - 2\pi_3 = 8$; da $x_3 \leq 0$ nasce $-\pi_2 \geq -9$.
- **Passo 3 — l'obiettivo**: il duale è un massimo, $\max\ 30\pi_1 + 100\pi_2 - 20\pi_3$.

$$\begin{array}{rrrrrrrr} \max & 30\,\pi_1 & + & 100\,\pi_2 & - & 20\,\pi_3 & & \\ \text{soggetto a} & \pi_1 & + & \pi_2 & + & \pi_3 & \leq & 5 \\ & \pi_1 & + & \pi_2 & - & 2\,\pi_3 & = & 8 \\ & - & & \pi_2 & & & \geq & -9 \\ & \pi_1 & & & & & \geq & 0 \\ & & & \pi_2 & & & \gtreqless & 0, \\ & & & & & \pi_3 & \leq & 0 \end{array}$$

2.1.2 Le condizioni di ottimalità: gli scarti complementari

A ogni vincolo del primale associamo il suo *scarto* $\bar{s}_i$, e a ogni variabile il suo *costo ridotto* $\bar{c}_j$, cioè lo scarto del corrispondente vincolo *duale*:

$$\bar{s}_i = \sum_{j \in N} a_{ij}\,\bar{x}_j - b_i, \quad \forall i \in M, \qquad \bar{c}_j = c_j - \sum_{i \in M} a_{ij}\,\bar{\pi}_i, \quad \forall j \in N.$$

Lo scarto misura quanto il vincolo i è lontano dall'essere attivo; il costo ridotto misura il coefficiente c_j al netto del valore delle risorse consumate da x_j, valutate con le variabili duali.

Condizioni di ottimalità dell'LP (scarti complementari)

Una coppia di soluzioni ammissibili, $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ per il primale (P) e $(\bar{\pi}_1, \bar{\pi}_2, \ldots, \bar{\pi}_m)$ per il duale (D), è ottima per entrambi i problemi *se e solo se*

$$\bar{\pi}_i \cdot \bar{s}_i = 0, \qquad \forall i \in M,$$
$$\bar{x}_j \cdot \bar{c}_j = 0, \qquad \forall j \in N.$$

Spiegazione riga per riga

In parole: **vincolo lasco** $\Rightarrow$ **duale a zero** ($\bar{s}_i \neq 0 \Rightarrow \bar{\pi}_i = 0$) e **variabile diversa da zero** $\Rightarrow$ **costo ridotto nullo** ($\bar{x}_j \neq 0 \Rightarrow \bar{c}_j = 0$). All'ottimo sono *implicazioni*, non equivalenze: per la degenerazione un vincolo attivo può avere duale nullo, e una variabile a zero costo ridotto nullo. Gli esempi di verifica più avanti controllano proprio queste condizioni.

Spiegazione riga per riga

Due letture da ricordare per il laboratorio:

- **Prezzo ombra** $\tilde{\pi}_i$: il valore marginale del termine noto b_i,

$$\tilde{\pi}_i = \frac{\partial \tilde{z}}{\partial b_i};$$

  vale per b_i nell'*intervallo di validità del termine noto* $[b_i^{\min}, b_i^{\max}]$: il più grande intervallo in cui, variando solo b_i, il prezzo ombra resta esattamente $\tilde{\pi}_i$ (oltre, cambia e va ricalcolato). Caso particolare: vincolo non attivo $\Rightarrow \tilde{\pi}_i = 0$.

- **Costo ridotto** $\bar{c}_j$ (lo scarto del vincolo duale, definito sopra): per una variabile a zero dice di quanto deve migliorare c_j perché convenga attivarla; vale per c_j nell'*intervallo di validità del coefficiente* $[c_j^{\min}, c_j^{\max}]$: il più grande intervallo in cui, variando solo c_j, la soluzione ottima non cambia (per una variabile a zero la soglia di convenienza è proprio uno dei due estremi). Caso particolare: variabile diversa da zero $\Rightarrow \bar{c}_j = 0$.

- **Attenzione alla degenerazione**: può esistere una variabile *a zero con costo ridotto nullo*, anche se non viene usata. Vale quindi solo l'implicazione "$\bar{c}_j \neq 0 \Rightarrow$ variabile su un suo bound", non il viceversa.

I segni dei prezzi ombra, spiegati bene

Il prezzo ombra è sempre, con un'unica convenzione,

$$\tilde{\pi}_i = \frac{\partial \tilde{z}}{\partial b_i},$$

la derivata del valore ottimo rispetto al termine noto del vincolo i: di quanto cambia $\tilde{z}$ se b_i aumenta di una (piccola) unità. Il segno si deduce sempre rispondendo a due domande:

- **aumentare** b_i **allarga o restringe la regione ammissibile?** Con un vincolo $\leq$ la allarga (il tetto si alza); con un vincolo $\geq$ la restringe (la richiesta cresce); con un'uguaglianza la sposta soltanto;

- **una regione più ampia come cambia l'ottimo?** Non può mai peggiorarlo: in un minimo il valore ottimo scende o resta uguale, in un massimo sale o resta uguale.

Combinando le due risposte:

Verso del vincolo	problema di minimo	problema di massimo
$\leq$ (b_i cresce $\Rightarrow$ regione più ampia)	$\tilde{\pi}_i \leq 0$	$\tilde{\pi}_i \geq 0$
$\geq$ (b_i cresce $\Rightarrow$ regione più stretta)	$\tilde{\pi}_i \geq 0$	$\tilde{\pi}_i \leq 0$
$=$	segno qualunque (duale libera)	

Due letture tipiche. In un *minimo*: un vincolo $\leq$ (una capacità) ha $\tilde{\pi}_i \leq 0$ — allentarlo riduce il costo, e $|\tilde{\pi}_i|$ è il risparmio per ogni unità di capacità in più; un vincolo $\geq$ (una richiesta da soddisfare) ha $\tilde{\pi}_i \geq 0$ — alzare la richiesta costa, e $\tilde{\pi}_i$ è il costo marginale. In un *massimo* le letture si ribaltano. In ogni caso un vincolo non attivo ha $\tilde{\pi}_i = 0$ per complementarietà.

Dualità su un LP 2×2, svolta per intero

Riprendiamo l'LP 2×2 e il suo duale, costruiti sopra, e verifichiamo le condizioni di ottimalità.

- **Passo 1 — risolvendo si ottiene:**

$$(\tilde{x}_1, \tilde{x}_2) = (30,\ 20), \qquad \tilde{z} = 1900, \qquad (\tilde{\pi}_1, \tilde{\pi}_2) = (14,\ 8), \qquad (\bar{c}_1, \bar{c}_2) = (0,\ 0),$$

  con intervalli di validità $[b_1^{\min}, b_1^{\max}] = [40, 240]$, $[b_2^{\min}, b_2^{\max}] = [30, 180]$, $[c_1^{\min}, c_1^{\max}] = [50/3,\ 100]$ e $[c_2^{\min}, c_2^{\max}] = [15,\ 90]$. Entrambi i vincoli sono attivi.

- **Passo 2 — verifica dei prezzi ombra.** Una unità in più di b_1 vale 14, una unità in più di b_2 vale 8. Verifichiamo la proprietà che li caratterizza: $\tilde{x}_1$ e $\tilde{x}_2$ sono diverse da zero, quindi per gli scarti complementari i loro vincoli duali devono essere attivi ($\bar{c}_1 = \bar{c}_2 = 0$):

$$\tilde{\pi}_1 + 2\tilde{\pi}_2 = 14 + 16 = 30 = c_1, \qquad 3\tilde{\pi}_1 + \tilde{\pi}_2 = 42 + 8 = 50 = c_2. \quad \checkmark$$

- **Passo 3 — verifica della dualità forte.**

$$\sum_{i \in M} b_i\, \tilde{\pi}_i = 90 \cdot 14 + 80 \cdot 8 = 1260 + 640 = 1900 = \tilde{z}. \quad \checkmark$$

- **Passo 4 — verifica dei costi ridotti.** Con la definizione $\bar{c}_j = c_j - \sum_{i \in M} a_{ij}\, \tilde{\pi}_i$:

$$\bar{c}_1 = 30 - (1 \cdot 14 + 2 \cdot 8) = 0, \qquad \bar{c}_2 = 50 - (3 \cdot 14 + 1 \cdot 8) = 0 :$$

  entrambi nulli, come per ogni variabile positiva all'ottimo: le due variabili ripagano esattamente le risorse che consumano. Un costo ridotto diverso da zero può averlo solo una variabile che sta su un proprio bound.

Tutti i casi in un solo LP: verifica delle condizioni

Riprendiamo l'LP "tutti i casi" e il suo duale (esempio 2.1.1) — i casi sono gli stessi che torneranno, uno alla volta, nei capitoli applicativi.

- **Passo 1 — risolvendo si ottiene:**

$$(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = (60,\ 40,\ 0), \qquad \tilde{z} = 620,$$

$$(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3) = (0,\ 6,\ -1), \qquad (\bar{c}_1, \bar{c}_2, \bar{c}_3) = (0,\ 0,\ -3),$$

con intervalli di validità dei termini noti $(-\infty, 100]$, $[30, +\infty)$ e $[-200, +\infty)$, e dei coefficienti $(-\infty, 8]$, $[5, 17]$ e $(-\infty, -6]$. All'ottimo sono attivi il secondo e il terzo vincolo, con $\tilde{x}_3 = 0$.

- **Passo 2 — verifica dei prezzi ombra con le regole della tabella.** $\tilde{\pi}_1 = 0$, coerente con la complementarietà: il primo vincolo è lasco ($\tilde{x}_1 + \tilde{x}_2 = 100 > 30$). $\tilde{\pi}_2 = 6$ (variabile duale libera, qui positiva). $\tilde{\pi}_3 = -1 \leq 0$, come dev'essere per un vincolo $\leq$ in un minimo. Verifichiamo che, come impongono gli scarti complementari, i vincoli duali di x_1 e x_2 (diverse da zero) siano attivi:

$$\underbrace{0 + 6 - 1}_{x_1} = 5 = c_1, \qquad \underbrace{0 + 6 - 2 \cdot (-1)}_{x_2} = 8 = c_2, \quad \checkmark$$

(x_2 è libera: il suo vincolo duale è di uguaglianza). Verifica per perturbazione:

$$b_2 = 101 \ \Rightarrow\ \tilde{z} = 626\ (+6), \qquad b_3 = -19 \ \Rightarrow\ \tilde{z} = 619\ (-1). \quad \checkmark$$

- **Passo 3 — verifica della dualità forte.**

$$\sum_{i \in M} b_i\, \tilde{\pi}_i = 30 \cdot 0 + 100 \cdot 6 + (-20) \cdot (-1) = 620 = \tilde{z}. \quad \checkmark$$

- **Passo 4 — verifica dei costi ridotti, uno per segno.** Con la definizione $\bar{c}_j = c_j - \sum_{i \in M} a_{ij}\, \tilde{\pi}_i$:

$$\bar{c}_1 = 5 - (0 + 6 - 1) = 0, \qquad \bar{c}_2 = 8 - (0 + 6 + 2) = 0 :$$

coerente con $\tilde{x}_1, \tilde{x}_2 \neq 0$. x_3 vale zero, che per una variabile ≤ 0 è il suo *bound superiore*:

$$\bar{c}_3 = -9 - (-1) \cdot 6 = -3,$$

cioè forzare $x_3 = -1$ peggiorerebbe l'ottimo di 3; x_3 si attiverebbe solo se il suo coefficiente salisse almeno a -6 (proprio l'estremo $c_3^{\max}$ dell'intervallo di validità). Verifica:

$$x_3 = -1 \ \Rightarrow\ \tilde{z} = 623\ (+3). \quad \checkmark$$

2.1.3 L'analisi di sensitività e i prezzi ombra

Risolvere un LP dà una soluzione ottima; l'**analisi di sensitività** risponde alla domanda successiva, quella che interessa davvero un decisore: *come cambia l'ottimo se cambiano i dati?* Negli LP gran parte della risposta arriva gratis insieme alla soluzione, sotto forma di quattro numeri per ogni vincolo e variabile:

- il **prezzo ombra** $\tilde{\pi}_i$ di un vincolo: di quanto migliora il valore ottimo se il termine noto del vincolo aumenta di una unità — "quanto pagherei per un'unità in più di quella risorsa";

- l'**intervallo di validità del termine noto** $[b_i^{\min}, b_i^{\max}]$: dove quel prezzo ombra resta esatto (fuori, va rifatta l'ottimizzazione);
- il **costo ridotto** $\bar{c}_j$ di una variabile a zero: di quanto deve migliorare il suo coefficiente in obiettivo perché convenga attivarla;
- l'**intervallo di validità del coefficiente** in obiettivo $[c_j^{\min}, c_j^{\max}]$: dove il coefficiente può variare senza che la soluzione ottima cambi. Nell'LP 2×2: c_1 può variare in $[50/3 \approx 16{,}7,\ 100]$ e c_2 in $[15, 90]$ senza spostare la soluzione $(30, 20)$.

Prezzi ombra in pratica, sull'LP 2×2

Riprendiamo l'LP 2×2: ottimo 1900 con duali $\tilde{\pi}_1 = 14$ e $\tilde{\pi}_2 = 8$.
Lettura. Una unità in più di b_1 fa crescere l'ottimo di 14; se procurarla costasse 12, converrebbe $(14 > 12)$.
Verifica per perturbazione. Rirrisolvendo:

$$b_1 = 91 \ \Rightarrow\ 1914 = 1900 + 14, \qquad b_1 = 100 \ \Rightarrow\ 2040 = 1900 + 10 \cdot 14 :$$

il prezzo ombra è ancora esatto perché 100 è dentro l'intervallo di validità $[40, 240]$.
Dove smette di valere. A $b_1 = 240$ il primo vincolo smette di essere la risorsa scarsa: oltre, ogni unità aggiuntiva vale meno di 14 e va ricalcolato tutto. Ecco perché ogni prezzo ombra si cita *insieme* al suo intervallo di validità.
Vincoli non attivi. Se fosse $b_1 = 300$ (vincolo non attivo all'ottimo), il suo prezzo ombra sarebbe 0: una risorsa che avanza non vale nulla al margine.

Costi ridotti in pratica: conviene una terza variabile?

Nell'LP 2×2 entrambe le variabili sono positive all'ottimo e i loro costi ridotti valgono 0: è sempre così per le variabili diverse da zero. Per vedere un costo ridotto all'opera serve una variabile a zero: aggiungiamo al modello una variabile $x_3 \geq 0$ con coefficiente $c_3 = 20$ in obiettivo e consumi $a_{13} = a_{23} = 1$ nei due vincoli.

- **Passo 1 — valutazione con i prezzi ombra.** Le risorse assorbite da un'unità di x_3 valgono, ai prezzi ombra,

  $$1 \cdot 14 + 1 \cdot 8 = 22 > 20 = c_3 :$$

  ogni unità di x_3 sottrarrebbe risorse a impieghi migliori.
- **Passo 2 — il costo ridotto.** La differenza è proprio il costo ridotto:

  $$\bar{c}_3 = 20 - 22 = -2.$$

  Risolvendo si conferma: la soluzione ottima resta $(30, 20, 0)$ con $\tilde{z} = 1900$, la variabile x_3 è a zero con $\bar{c}_3 = -2$ e $c_3^{\max} = 22$: il coefficiente c_3 dovrebbe salire almeno a 22 perché convenga attivarla.
- **Passo 3 — controprova.** Con $c_3 = 23 > 22$ la soluzione cambia davvero, e non di poco: la nuova soluzione ottima è $(0, 5, 75)$ con valore 1975. Il costo ridotto dice *quando* una variabile a zero entra in soluzione, non *come* cambia la soluzione: quella va ricalcolata.

2.2 Il protocollo di analisi di sensitività

L'analisi di sensitività è parte del modello, non un'appendice. In ogni esercitazione seguiamo lo stesso protocollo in sei passi:

1. Scenario base	risolvere, verificare l'ammissibilità, identificare i vincoli attivi.
2. One-at-a-time	variare un parametro chiave su una griglia; grafico valore/decisioni.
3. Prezzi ombra e costi ridotti	confrontare la previsione marginale del duale con una ri-ottimizzazione dopo una piccola perturbazione; leggere il costo ridotto $\bar{c}_j$ di ogni variabile a zero: la soglia oltre la quale conviene attivarla.
4. Scenari	costruire almeno tre scenari coerenti (pessimistico, centrale, ottimistico).
5. Trade-off	variare un peso o una soglia per tracciare una frontiera (costo-servizio, rischio-rendimento, costo-emissioni).
6. Stabilità	segnalare cambi bruschi del mix ottimo e parametri stimati fragili.

2.3 In sintesi

In sintesi

- Ogni vincolo genera una variabile duale e ogni variabile un vincolo duale; la dualità debole dà un bound, quella forte valori uguali all'ottimo.
- Scarti complementari: un vincolo non attivo ha duale nullo, una variabile non nulla ha costo ridotto nullo.
- Il prezzo ombra $\tilde{\pi}_i = \partial \tilde{z}/\partial b_i$ vale solo dentro il suo intervallo; un numero si riporta sempre con il suo intervallo.
- Il costo ridotto dice di quanto deve migliorare c_j perché convenga usare una variabile ferma a zero.
- Il protocollo: caso base, one-at-a-time, prezzi ombra verificati per perturbazione, scenari, trade-off, stabilità.

Esercizio 2.1 — dualità

Nell'LP 2×2 dell'esempio è possibile aumentare b_1 di 10 unità pagando 12 per unità. Conviene? Fino a che prezzo unitario conviene? E fino a quale valore di b_1 vale il prezzo ombra di 14? (Suggerimento: ricalcolare il vertice ottimo quando il primo vincolo si sposta a $90 + \Delta$ e trovare il Δ oltre il quale la soluzione cambia struttura.)

Capitolo 3

Il solver: costruire, risolvere, interpretare

Prerequisiti

- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: duali, costi ridotti, scarti e intervalli di validità, che il solver restituisce come attributi.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- costruire un LP in `gurobipy` in cinque passi;
- dichiarare le variabili libere, ricordando il lower bound predefinito 0;
- controllare `Status` prima di leggere `ObjVal`, `X` e `Slack`;
- leggere `Pi`, `RC`, `SARHS` e `SAObj` come le grandezze del capitolo dei richiami;
- riconoscere gli errori tipici, dal modello non ammissibile al termine noto che contiene costanti.

Questo capitolo è la cassetta degli attrezzi del laboratorio: come si costruisce un modello con `gurobipy`, come si fa girare, come si recupera la soluzione e — la parte più importante — come si interpreta l'output. Gli esempi sono generici e usano gli stessi LP del capitolo 2: insieme, i due capitoli danno le basi per capire tutti i modelli e gli esempi dei capitoli successivi. Una versione autonoma di questa guida è nel file `GUIDA_GUROBI.md` nella cartella del corso. Due sottocapitoli: i *comandi generali* (costruire, risolvere, leggere la soluzione), poi i comandi specifici per i *modelli lineari* (duali, costi ridotti, intervalli di validità).

3.1 I comandi generali

3.1.1 Installazione e licenza

```
python3 -m pip install gurobipy
```

Il pacchetto include una licenza dimostrativa (fino a 2000 variabili e 2000 vincoli), sufficiente per quasi tutti i modelli della dispensa. Per la licenza accademica completa e gratuita: registrarsi su `https://portal.gurobi.com` con l'email istituzionale, richiedere una *Named-User*

Academic License ed eseguire il comando `grbgetkey` indicato dal portale (sotto rete di ateneo o VPN).

3.1.2 I cinque passi per costruire un modello

Riprendiamo l'LP 2×2 del capitolo 2: $\max 30x_1+50x_2$ con $x_1+3x_2 \le 90$ (vincolo 1), $2x_1+x_2 \le 80$ (vincolo 2) e $x_1, x_2 \ge 0$.

```
import gurobipy as gp
from gurobipy import GRB

# Passo 1: il contenitore del modello
m = gp.Model("lp_2x2")

# Passo 2: le variabili decisionali (default: continue, >= 0)
x1 = m.addVar(name="x1")
x2 = m.addVar(name="x2")

# Passo 3: i vincoli (conservare l'oggetto per leggere poi il duale)
v1 = m.addConstr(x1 + 3*x2 <= 90, name="vincolo1")
v2 = m.addConstr(2*x1 + x2 <= 80, name="vincolo2")

# Passo 4: la funzione obiettivo
m.setObjective(30*x1 + 50*x2, GRB.MAXIMIZE)

# Passo 5 (debug): scrivere il modello in formato leggibile
m.write("lp_2x2.lp")
```

Le tre regole d'oro di `gurobipy`

1. **Variabili indicizzate** con `addVars`:
   `x = m.addVars(I, T, name="x")` crea la famiglia `x[i,t]` sugli insiemi di indici `I` e `T`; le somme si scrivono `gp.quicksum(...)` o `x.sum(i, "*")`.
2. **Variabili libere** (che possono essere negative, come l'intercetta b della regressione o la soglia η del CVaR): `m.addVar(lb=-GRB.INFINITY)`. Dimenticare `lb` è l'errore più frequente del laboratorio: il default è ≥ 0!
3. **Famiglie di vincoli** con generatore: `m.addConstrs((x.sum("*", t) <= b[t] for t in T), name="cap")`.

3.1.3 Far girare il modello e leggere il log

```
m.Params.OutputFlag = 0     # 0 = silenzioso, 1 = log a video
m.Params.TimeLimit  = 60    # secondi
m.optimize()
```

Il log del solver, quando attivo, va letto in tre punti:

```
Optimize a model with 2 rows, 2 columns and 4 nonzeros   <- dimensioni attese?
Coefficient statistics: Matrix range [1e+00, 3e+00]      <- dati ben scalati?
Optimal objective  1.900000000e+03                       <- esito e valore
```

Se i range dei coefficienti spaziano da 10^{-6} a 10^{9} il modello è mal scalato: cambiare unità di misura (migliaia di euro, MWh, ...) prima di cercare altri errori.

3.1.4 Recuperare la soluzione

Controllare sempre lo stato prima di leggere qualunque numero:

```
if m.Status == GRB.OPTIMAL:
    print(m.ObjVal, x1.X, x2.X)
elif m.Status == GRB.INFEASIBLE:
    m.computeIIS(); m.write("conflitto.ilp")    # quali vincoli si
        contraddicono
elif m.Status == GRB.UNBOUNDED:
    pass  # quasi sempre manca un vincolo o un bound
```

Attributo	Oggetto	Significato
`Status`	modello	`OPTIMAL`=2, `INFEASIBLE`=3, `UNBOUNDED`=5, `TIME_LIMIT`=9
`ObjVal`	modello	valore ottimo dell'obiettivo
`X`	variabile	valore della variabile all'ottimo
`Slack`	vincolo	scarto: 0 ⇒ vincolo attivo

Gli attributi specifici degli LP (prezzi ombra, costi ridotti, intervalli di validità) sono nel sottocapitolo che segue.

3.2 I comandi per i modelli lineari

3.2.1 Prezzi ombra, costi ridotti e intervalli: gli attributi

Il ponte con la notazione del capitolo 2

Gli attributi del solver sono esattamente le quantità teoriche dei richiami:

Teoria (cap. 2)	Attributo Gurobi	Oggetto
prezzo ombra $\tilde{\pi}_i$	`Pi`	vincolo
costo ridotto $\bar{c}_j$	`RC`	variabile
scarto $\bar{s}_i$	`Slack`	vincolo
intervallo $[b_i^{\min}, b_i^{\max}]$	`SARHSLow`–`SARHSUp`	vincolo
intervallo $[c_j^{\min}, c_j^{\max}]$	`SAObjLow`–`SAObjUp`	variabile

Se un vincolo contiene una costante a sinistra, come `quicksum(x[v,t] ...) + base[t] <= C`, Gurobi la sposta nel termine noto: il RHS memorizzato è `C - base[t]`. Chi in un'analisi di sensitività scrive `constr.RHS = nuovaC` sta in realtà imponendo `quicksum(x) <= nuovaC`, cioè un'altra cosa. La forma corretta è `constr.RHS = nuovaC - base[t]`. Questo errore ci è capitato davvero preparando il caso di studio della ricarica dei veicoli elettrici, che lo discute.

3.2.2 Interpretare l'output: la checklist

1. Lo stato è `OPTIMAL`? Se no, fermarsi e diagnosticare (`computeIIS` per l'inammissibilità, file `.lp` per l'illimitatezza).
2. Il valore ottimo ha l'ordine di grandezza atteso?

3. Quali vincoli sono attivi (`Slack` $= 0$)? Sono quelli previsti?
4. Quanto valgono i prezzi ombra? Quale risorsa conviene potenziare per prima? (Ricordare i segni: in un problema di minimo, un vincolo $\leq$ ha `Pi` ≤ 0.)
5. Ci sono variabili a zero? Leggere il loro costo ridotto (`RC`): dice di quanto deve migliorare il coefficiente in obiettivo perché convenga attivarle (paragrafo 2.1.3).
6. La soluzione è stabile? Rieseguire con dati perturbati del 5%.
7. Tradurre tutto in una raccomandazione manageriale di tre righe.

3.2.3 Due letture complete

Lettura completa dell'esempio 2×2

Eseguendo il codice della sezione precedente:

```
Status: 2 (OPTIMAL)       ObjVal: 1900.0
x1 = 30.0    RC = 0.0    SAObj = [16.7, 100.0]
x2 = 20.0    RC = 0.0    SAObj = [15.0, 90.0]
vincolo1: Slack = 0.0    Pi = 14.0    SARHS = [40.0, 240.0]
vincolo2: Slack = 0.0    Pi = 8.0     SARHS = [30.0, 180.0]
```

Lettura: entrambi i vincoli sono attivi (saturi); una unità in più di b_1 vale 14, e la valutazione resta valida finché b_1 resta in $[40, 240]$; oltre, va ricalcolato tutto. Se si potesse aumentare un solo termine noto, meglio b_1 ($14 > 8$ per unità, a parità di costo). I costi ridotti (attributo `x1.RC`) sono nulli perché entrambe le variabili sono in base — attenzione: può esistere anche una variabile *in base a valore zero* (degenerazione), quindi `RC` $= 0$ da solo non dice che la variabile è usata. Aggiungendo la terza variabile del capitolo 2 il solver darebbe $x_3 = 0$ con `RC` $= -2$: il suo coefficiente deve salire di almeno 2 (cioè a 22) perché entri in soluzione, con range di validità `SAObj` $= (-\infty, 22]$.

Nel caso generale: che cosa restituisce il solver e con che segno

Prezzi ombra (attributo `Pi`, uno per vincolo): sempre la "derivata del valore ottimo rispetto al termine noto". Il segno si deduce in due mosse: aumentare b_i allarga la regione ammissibile (vincolo $\leq$) o la restringe (vincolo $\geq$); e una regione più ampia non può mai peggiorare l'ottimo. Quindi:

- in un problema di *massimo*: vincolo $\leq$ di risorsa $\Rightarrow$ `Pi` ≥ 0 (più risorsa, più profitto); vincolo $\geq \Rightarrow$ `Pi` ≤ 0;
- in un problema di *minimo*: vincolo $\geq$ di domanda $\Rightarrow$ `Pi` ≥ 0 (più domanda, più costo); vincolo $\leq$ di capacità $\Rightarrow$ `Pi` ≤ 0 (più capacità, meno costo);
- vincolo di *uguaglianza* $\Rightarrow$ segno qualunque, come da regole della dualità. Vincolo non attivo $\Rightarrow$ `Pi` $= 0$; validità nel range `SARHSLow`–`SARHSUp`.

Costi ridotti (attributo `RC`, uno per variabile):

- variabile *in base*: `RC` $= 0$ — di norma è strettamente tra i suoi bound, ma può anche trovarsi su un bound con `RC` $= 0$ (base *degenere*);

- variabile al *bound inferiore* (tipicamente a zero): `RC` ≥ 0 in un minimo (di quanto deve scendere il suo costo perché convenga usarla), `RC` ≤ 0 in un massimo (di quanto deve salire il suo margine); validità nel range `SAObjLow`–`SAObjUp`;
- variabile al *bound superiore*: `RC` è il prezzo ombra del bound — in un modello di flusso di minimo costo un arco saturo ha `RC` < 0: un'unità di capacità in più su quell'arco fa risparmiare $|$`RC`$|$.

Il ciclo completo, per *tutti* i vincoli e le variabili (solo LP):

```
for v in m.getConstrs():
    print(v.ConstrName, v.Pi, v.SARHSLow, v.SARHSUp)
for v in m.getVars():
    print(v.VarName, v.X, v.RC, v.SAObjLow, v.SAObjUp)
```

L'esempio completo: tutti i casi, dal modello alla lettura

Implementiamo l'LP "tutti i casi" del capitolo 2 (tre versi di vincolo, tre segni di variabile) e leggiamo *tutto* quello che il solver restituisce:

```
m = gp.Model("tutti_i_casi")
x1 = m.addVar(name="x1")                          # x1 >= 0 (default)
x2 = m.addVar(lb=-GRB.INFINITY, name="x2")        # x2 libera
x3 = m.addVar(lb=-GRB.INFINITY, ub=0, name="x3")  # x3 <= 0
v1 = m.addConstr(x1 + x2 >= 30,        name="vincolo1")  # verso >=
v2 = m.addConstr(x1 + x2 - x3 == 100,  name="vincolo2")  # verso  =
v3 = m.addConstr(x1 - 2*x2 <= -20,     name="vincolo3")  # verso <=
m.setObjective(5*x1 + 8*x2 - 9*x3, GRB.MINIMIZE)
m.optimize()
for v in m.getVars():
    print(v.VarName, v.X, v.RC, v.SAObjLow, v.SAObjUp)
for c in m.getConstrs():
    print(c.ConstrName, c.Slack, c.Pi, c.SARHSLow, c.SARHSUp)
```

```
Status: 2 (OPTIMAL)    ObjVal: 620.0
x1: X =  60.0  RC =  0.0  SAObj = [-inf, 8.0]
x2: X =  40.0  RC =  0.0  SAObj = [5.0, 17.0]
x3: X =   0.0  RC = -3.0  SAObj = [-inf, -6.0]
vincolo1: Slack = -70.0  Pi =  0.0  SARHS = [-inf, 100.0]
vincolo2: Slack =   0.0  Pi =  6.0  SARHS = [30.0, inf]
vincolo3: Slack =   0.0  Pi = -1.0  SARHS = [-200.0, inf]
```

Qui compaiono tutti i casi della spiegazione precedente: il vincolo non attivo con `Pi` $= 0$ (`vincolo1`, `Slack` $\neq 0$); l'uguaglianza con duale positivo ($+6$); il $\leq$ in un minimo con duale negativo (-1); due variabili in base con `RC` $= 0$; e la variabile x_3 ferma al suo bound superiore (0) con `RC` $= -3$ e soglia di convenienza `SAObjUp` $= -6$. La verifica a mano di ognuno di questi numeri, con le regole della dualità e per perturbazione, è nell'esempio "tutti i casi" del capitolo 2.

3.3 In sintesi

In sintesi

- Cinque passi: `Model`, `addVar(s)`, `addConstr(s)`, `setObjective`, `write`.
- Le variabili sono ≥ 0 per default: una variabile libera richiede `lb=-GRB.INFINITY`.
- Prima di tutto `Status`; solo dopo `ObjVal`, `X` e `Slack`.
- `Pi`, `RC`, `Slack`, `SARHS` e `SAObj` sono le grandezze teoriche dei richiami; il termine noto memorizzato include le costanti portate dal membro sinistro.
- `RC` $= 0$ non basta per dire che una variabile è usata: la degenerazione è sempre possibile.

Esercizio 3.1

Implementare l'LP 2×2, stampare `Pi`, `Slack`, `SARHSLow/Up` di entrambi i vincoli e verificare il prezzo ombra del primo vincolo ri-ottimizzando con $b_1 : 90 \to 91$.

Esercizio 3.2

Togliere il secondo vincolo e osservare l'esito. Poi cambiare l'obiettivo in minimizzazione senza altri cambiamenti: perché il risultato è $x = 0$?

Esercizio 3.3

Costruire apposta un modello inammissibile (due vincoli incompatibili, ad esempio $x_1 + x_2 \geq 10$ e $x_1 + x_2 \leq 8$) e usare `computeIIS` per individuare il conflitto.

Parte II

Modelli deterministici

Capitolo 4

Produzione e scorte multiperiodali

Classe: LP **Script:** `python/lab04_produzione.py`

Pagina online

Prerequisiti

- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: i prezzi ombra, i loro intervalli di validità e il protocollo di sensitività.
- Da questo modulo, Capitolo 3 «Il solver: costruire, risolvere, interpretare»: la costruzione di un LP in `gurobipy` e la lettura di `Pi` e `SARHS`.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- formulare un LP multiperiodale con vincoli di bilancio delle scorte;
- leggere e verificare i prezzi ombra della capacità e i loro intervalli di validità;
- costruire la curva del valore della capacità;
- spiegare come le scorte spostano capacità dai mesi scarichi ai mesi di picco.

4.1 Motivazione gestionale

Un'azienda decide quanto produrre oggi e quanto conservare a scorta per la domanda futura. Produrre in anticipo costa giacenza; produrre all'ultimo momento rischia di sbattere contro il limite di capacità proprio nei mesi di picco. Il modello bilancia margini, capacità, scorte e stabilità del piano — ed è il punto di partenza ideale del laboratorio perché ogni suo numero ha un'interpretazione economica immediata.

In pratica questo problema si presenta ogni volta che un'azienda pianifica la produzione su più periodi con risorse condivise: piani principali di produzione (MPS), pianificazione aggregata, contratti di fornitura con profili mensili. Il nucleo della difficoltà è il *legame intertemporale*: la decisione di oggi (produrre in anticipo o no) cambia i vincoli di domani.

4.2 Costruzione guidata del modello

Il problema a parole. *Decidiamo* quante unità di ogni prodotto produrre in ogni mese e quante tenerne a scorta. *L'obiettivo* è minimizzare la somma dei costi di produzione e di giacenza sull'intero orizzonte. *I vincoli* sono due: la contabilità di magazzino deve tornare in ogni mese (ciò che entra uguaglia ciò che esce, e tutta la domanda va servita) e le ore macchina usate in un mese non possono superare quelle disponibili.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di prodotti; i prodotti sono indicizzati da $i \in \{1, 2, \dots, n\}$
m	$\in \mathbb{Z}_{\geq 1}$	numero di mesi dell'orizzonte; i mesi sono indicizzati da $t \in \{1, 2, \dots, m\}$
d_{it}	$\in \mathbb{Q}_{\geq 0}$	domanda del prodotto i nel mese t (unità)
c_{it}	$\in \mathbb{Q}_{>0}$	costo di produzione di un'unità di i nel mese t (€)
h_{it}	$\in \mathbb{Q}_{>0}$	costo di giacenza di un'unità di i per il mese t (€)
a_i	$\in \mathbb{Q}_{>0}$	ore macchina per produrre un'unità del prodotto i
b_t	$\in \mathbb{Q}_{\geq 0}$	ore macchina disponibili nel mese t
s_{i0}	$\in \mathbb{Q}_{\geq 0}$	scorta iniziale del prodotto i (unità)

Variabili decisionali

Introduciamo le seguenti $2\,n\,m$ variabili non negative:

$$\begin{cases} x_{it} = \text{unità del prodotto } i \text{ prodotte nel mese } t \\ s_{it} = \text{unità del prodotto } i \text{ a scorta alla fine del mese } t \end{cases}$$

$$\forall i \in \{1, 2, \dots, n\},\ \forall t \in \{1, 2, \dots, m\}.$$

Usando queste variabili, un modello LP per il problema è il seguente:

Modello LP di produzione e scorte a costo minimo

$$\min \quad \sum_{i=1}^{n} \sum_{t=1}^{m} \left(c_{it}\, x_{it} + h_{it}\, s_{it} \right) \tag{4.1a}$$

$$\text{soggetto a} \quad s_{i,t-1} + x_{it} = d_{it} + s_{it}, \quad \forall i \in \{1, 2, \dots, n\},\ \forall t \in \{1, 2, \dots, m\}, \tag{4.1b}$$

$$\sum_{i=1}^{n} a_i\, x_{it} \leq b_t, \qquad \forall t \in \{1, 2, \dots, m\}, \tag{4.1c}$$

$$x_{it} \geq 0, \qquad \forall i \in \{1, 2, \dots, n\},\ \forall t \in \{1, 2, \dots, m\}, \tag{4.1d}$$

$$s_{it} \geq 0, \qquad \forall i \in \{1, 2, \dots, n\},\ \forall t \in \{1, 2, \dots, m\}. \tag{4.1e}$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (4.1a) minimizza il costo totale del piano, somma dei costi di produzione $c_{it}\, x_{it}$ e di giacenza $h_{it}\, s_{it}$ su tutti i prodotti e i mesi;

- i vincoli lineari (4.1b) sono i bilanci delle scorte: per ogni prodotto e ogni mese, ciò che entra (scorta precedente più produzione) è uguale a ciò che esce (domanda servita più scorta successiva); per $t = 1$ si usa la scorta iniziale s_{i0} ($n\,m$ vincoli lineari);
- i vincoli lineari (4.1c) assicurano che, in ogni mese, le ore macchina usate da tutti i prodotti non superino quelle disponibili (m vincoli lineari);
- i vincoli (4.1d) e (4.1e) definiscono le variabili del modello.

Spiegazione riga per riga

Perché il bilancio (4.1b) è il cuore del modello. È il vincolo che *lega i mesi tra loro*: senza di esso il problema si spezzerebbe in m problemi indipendenti e le scorte non avrebbero senso.
Cosa succede se... togliamo il bilancio? Il modello produce zero (costo minimo!) perché nulla lo obbliga a servire la domanda. Se la domanda totale supera la capacità cumulata, il modello diventa inammissibile: in tal caso o si consente la domanda persa con una variabile $u_{it} \geq 0$ e una penalità π_{it} in obiettivo, oppure si rivede la capacità.

4.3 Esempio svolto a mano

Un prodotto, due mesi

Domanda $d_1 = 60$, $d_2 = 100$; capacità $b_1 = b_2 = 80$ unità/mese ($a = 1$); costo di produzione $c = 10$ €, giacenza $h = 1$ €/unità/mese, scorta iniziale nulla.

- **Passo 1 — è ammissibile?** Domanda totale 160 $\leq$ capacità totale 160: sì, ma solo producendo al massimo in entrambi i mesi.
- **Passo 2 — bilanci.** $x_1 = 60 + s_1$ e $s_1 + x_2 = 100$. Con $x_2 \leq 80$ serve $s_1 \geq 20$, quindi $x_1 \geq 80$: ma $x_1 \leq 80$, dunque $x_1 = 80$, $s_1 = 20$, $x_2 = 80$.
- **Passo 3 — costo.** $10 \cdot (80 + 80) + 1 \cdot 20 = 1620$ €. La soluzione è *forzata*: entrambi i vincoli di capacità sono attivi.
- **Passo 4 — prezzo ombra a mano.** Con un'ora in più nel mese 2 ($b_2 = 81$): $x_2 = 81$, $s_1 = 19$, $x_1 = 79$: il costo scende di 1 € (un'unità non sta più a scorta un mese). Il duale della capacità del mese 2 è dunque -1 €/unità. Con un'ora in più nel mese 1 invece non cambia nulla (il mese 1 produce già il necessario): duale 0. Questo meccanismo — *la capacità di un mese vale quanto giacenza fa risparmiare* — riappare identico nel caso di studio.

4.4 Caso di studio

Tre prodotti ($n = 3$) su sei mesi ($m = 6$), con domanda stagionale crescente (picco nei mesi 4–5), un'unica risorsa condivisa e questi parametri (file CSV `produzione_domanda` e `produzione_capacita` nella cartella `dati/`):

Prodotto	c_i (€/u)	h_i (€/u/mese)	a_i (ore/u)	s_{i0}
1	12,00	0,80	0,9	30
2	18,00	1,20	1,4	20
3	25,00	1,60	2,1	10

mese	1	2	3	4	5	6
b_t (ore)	420	420	460	460	460	420
$d_{1,t}$	110	130	150	190	210	160
$d_{2,t}$	70	80	110	140	150	100
$d_{3,t}$	50	55	60	80	95	70

4.5 Implementazione

Il cuore dello script (versione integrale in `lab04_produzione.py`):

```
m = gp.Model("produzione_scorte")
x = m.addVars(prodotti, mesi, name="x")    # quantità prodotta
s = m.addVars(prodotti, mesi, name="s")    # scorta a fine mese

# bilancio: scorta entrante + produzione
#           = domanda + scorta uscente
m.addConstrs(
    ((s0[i] if t == 1 else s[i, t-1]) + x[i, t]
     == domanda[i, t] + s[i, t]
     for i in prodotti for t in mesi), name="bilancio")

# capacità condivisa (conservo i vincoli per i duali)
v_cap = m.addConstrs(
    (gp.quicksum(a[i] * x[i, t] for i in prodotti)
     <= capacita[t] for t in mesi), name="capacita")

m.setObjective(
    gp.quicksum(c[i]*x[i, t] + h[i]*s[i, t]
                for i in prodotti for t in mesi),
    GRB.MINIMIZE)
m.optimize()
```

4.6 Risultati

```
Costo totale ottimo: 32.889,33 EUR

Piano di produzione (unita'/mese):
mese       1      2      3      4      5      6
prod.1  80.0  130.0  150.0  190.0  210.0  160.0
prod.2  50.0   80.0  110.0  140.0  150.0  100.0
prod.3  89.5   91.0   81.4   44.3   29.0   64.8

Scorte a fine mese:
mese       1     2      3     4    5    6
prod.1   0.0   0.0    0.0   0.0  0.0  0.0
prod.2   0.0   0.0    0.0   0.0  0.0  0.0
prod.3  49.5  85.5  106.9  71.2  5.2  0.0
```

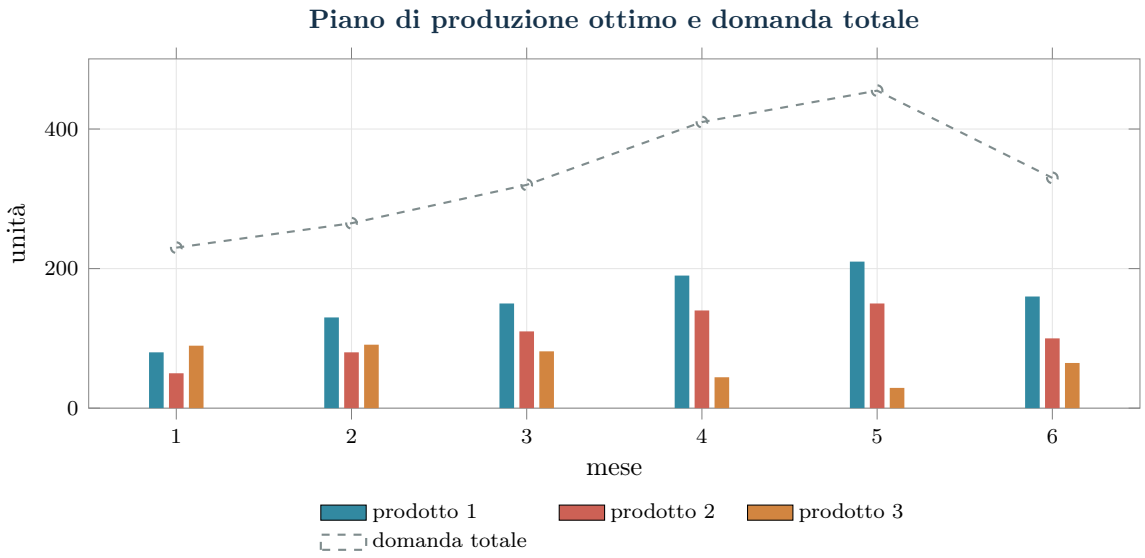


Figura 4.1: Piano di produzione ottimo dell'LP: barre = unità prodotte dei tre prodotti in ciascun mese; linea tratteggiata = domanda totale del mese. Nei mesi di picco (4–5) la produzione totale resta sotto la domanda: la differenza è coperta dalle scorte del prodotto 3 accumulate nei primi mesi.

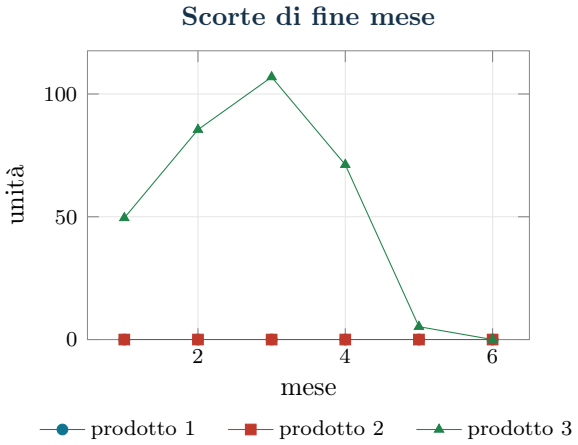


Figura 4.2: Scorte di fine mese per prodotto — solo il prodotto 3 usa il magazzino (strategia di pre-build), e le scorte si azzerano nel mese finale.

Insight manageriale

Il solver ha scoperto da solo la strategia del *pre-build*: i prodotti 1 e 2 (margini di giacenza alti rispetto alle ore assorbite) vengono prodotti su domanda, mentre il prodotto 3 — il più oneroso in ore macchina — viene prodotto in anticipo nei mesi 1–3 e accumulato (fino a 107 unità a scorta) per liberare capacità nei mesi di picco. Le scorte tornano a zero nel mese finale: una scorta terminale positiva sarebbe denaro sprecato, e il modello lo sa.

4.7 Analisi di sensitività

Prezzi ombra della capacità

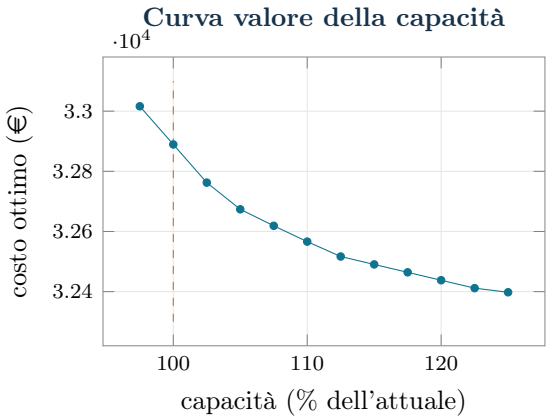
```
mese 1: uso 330.0/420  | Pi =  0.000 EUR/ora | valido per b_t in [330.0, +inf]
mese 2: uso 420.0/420  | Pi = -0.762 EUR/ora | valido per b_t in [330.0, 524.0]
mese 3: uso 460.0/460  | Pi = -1.524 EUR/ora | valido per b_t in [370.0, 564.0]
mese 4: uso 460.0/460  | Pi = -2.286 EUR/ora | valido per b_t in [370.0, 564.0]
mese 5: uso 460.0/460  | Pi = -3.048 EUR/ora | valido per b_t in [399.0, 564.0]
mese 6: uso 420.0/420  | Pi = -3.810 EUR/ora | valido per b_t in [330.0, 431.0]

Verifica: +1 ora nel mese 6 -> costo da 32889.33 a 32885.52 (delta -3.810)
```

Da dove viene il numero 0,762

I duali crescono di esattamente 0,762 € al mese. Non è un caso: $0{,}762 = h_3/a_3 = 1{,}6/2{,}1$. Un'ora in più nel mese t permette di produrre $1/a_3$ unità del prodotto 3 *un mese più tardi*, risparmiando un mese di giacenza: $h_3 \cdot (1/a_3)$. Il duale del mese 6 vale $5 \times 0{,}762 = 3{,}81$ perché un'ora lì evita cinque mesi di giacenza a un'unità prodotta nel mese 1. Quando il solver e un ragionamento economico elementare danno lo stesso numero, il modello è probabilmente giusto.

Il valore della capacità



La curva è **convessa e decrescente**: le prime ore extra valgono molto, poi sempre meno. Sotto il 97% dell'attuale capacità il modello diventa *inammissibile*: la domanda non è più servibile per intero — informazione gestionale preziosa quanto l'ottimo stesso.

Figura 4.3: Costo totale ottimo in funzione della capacità disponibile (in % dell'attuale; la linea tratteggiata arancione segna la capacità corrente): curva convessa e decrescente, interrotta a sinistra dove il problema diventa inammissibile.

Insight manageriale

Raccomandazione tipo da relazione di laboratorio: "Conviene acquistare ore straordinarie nei mesi 5–6 fino a 3–3,8 €/ora sotto il prezzo interno; sotto il 97% della capacità attuale il servizio completo non è garantibile."

Un piano *regolare*, che penalizza le variazioni della produzione da un mese all'altro, richiede un termine quadratico nell'obiettivo e non è più un LP: è la variante quadratica di questo modello, nel modulo *Modellazione PNL*, pagina «Produzione con smoothing quadratico».

4.8 In sintesi

In sintesi

- Il bilancio delle scorte lega i mesi: il problema non si risolve un mese alla volta.
- Le scorte spostano capacità dai mesi scarichi ai mesi di picco; svuotare il magazzino alla fine è sempre ottimo.
- Il prezzo ombra di una capacità si verifica aggiungendo un'ora e riottimizzando; una capacità non attiva ha prezzo ombra nullo.
- Un'ora di capacità in più vale la giacenza che fa risparmiare; la curva del valore della capacità è convessa e decrescente.

4.9 Esercizi

Esercizio 4.1 — verifica

Verificare per perturbazione i duali dei mesi 3 e 4 (aggiungere un'ora e riottimizzare). Perché il duale del mese 1 è nullo?

Esercizio 4.2 — domanda persa

Aggiungere la variabile u_{it} (domanda non servita) con penalità $\pi = 40$ €/unità e ridurre la capacità al 90%: quanta domanda si perde, di quale prodotto, in quali mesi? Interpretare la scelta del modello.

Esercizio 4.3 — massimo profitto

Trasformare il modello in massimizzazione del profitto con ricavi $p = (20, 30, 42)$ € e servizio facoltativo ($y_{it} \le d_{it}$). Quale domanda conviene *non* servire, e perché?

Esercizio 4.4 — emissioni

Ogni unità del prodotto 3 emette 2,5 kg di CO_2, i prodotti 1 e 2 un kg. Aggiungere il termine $\tau \sum_{i=1}^{n} \sum_{t=1}^{m} e_i\, x_{it}$ all'obiettivo e tracciare la frontiera costo-emissioni per $\tau \in [0, 10]$ €/kg.

Esercizio 4.5 — scorta di sicurezza

Imporre $s_{it} \ge 0{,}1\, d_{i,t+1}$ (scorta di sicurezza del 10%). Quanto costa questa politica? Quale vincolo di capacità diventa più critico?

Capitolo 5

Ricarica intelligente di veicoli elettrici

Classe: LP **Script:** `python/lab05_ricarica_ev.py`

Pagina online

Prerequisiti

- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: i prezzi ombra e il protocollo di sensitività.
- Da questo modulo, Capitolo 3 «Il solver: costruire, risolvere, interpretare»: i bound delle variabili in `gurobipy` e il termine noto memorizzato dal solver.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- modellare decisioni su un orizzonte orario con finestre di disponibilità;
- leggere i duali come prezzi delle ore marginali;
- gestire il compromesso costo-picco con un obiettivo pesato;
- diagnosticare una vera insidia dell'analisi di sensitività: il termine noto con le costanti.

5.1 Motivazione gestionale

Una flotta aziendale deve arrivare carica al mattino sfruttando le ore notturne a prezzo basso — ma se tutti i veicoli caricano insieme nelle ore economiche, il prelievo di punta esplode e con esso i costi di potenza e il rischio di superare la capacità del contatore. Il modello decide *quanta potenza* dare a ciascun veicolo in ciascuna ora: una decisione naturalmente continua, senza alcun bisogno di variabili di accensione/spegnimento.

La ricarica intelligente (*smart charging*) è uno dei problemi operativi nuovi creati dalla transizione energetica: flotte elettriche, prezzi orari dell'energia, limiti di potenza dei contatori. È anche un esempio perfetto di problema che *sembra* richiedere variabili di accensione/spegnimento e invece è un LP puro, perché la decisione vera — quanta potenza — è continua.

5.2 Costruzione guidata del modello

Il problema a parole. *Decidiamo* quanta potenza erogare a ogni veicolo in ogni ora. *L'obiettivo* è minimizzare la spesa energetica della notte. *I vincoli*: ogni veicolo deve ricevere tutta l'energia che gli serve entro la partenza, può caricare solo quando è collegato e mai oltre la potenza del suo caricatore, e il prelievo complessivo (ricarica più edificio) non può mai superare la potenza del contatore.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di veicoli, indicizzati da $v \in \{1, 2, \ldots, n\}$
m	$\in \mathbb{Z}_{\geq 1}$	numero di intervalli orari, indicizzati da $t \in \{1, 2, \ldots, m\}$, di durata Δt (qui $\Delta t = 1$ h)
π_t	$\in \mathbb{Q}_{\geq 0}$	prezzo dell'energia nell'ora t (€/kWh)
a_{vt}	$\in \{0, 1\}$	*dato* osservato: 1 se il veicolo v è collegato nell'ora t, 0 altrimenti
e_v	$\in \mathbb{Q}_{>0}$	energia richiesta dal veicolo v entro la partenza (kWh)
$\bar{p}_v$	$\in \mathbb{Q}_{>0}$	potenza massima di ricarica del veicolo v (kW)
b_t	$\in \mathbb{Q}_{\geq 0}$	carico di base dell'edificio nell'ora t (kW)
k	$\in \mathbb{Q}_{>0}$	potenza massima prelevabile dal contatore (kW)
η	$\in \mathbb{Q},\ 0 < \eta \leq 1$	rendimento di carica

Variabili decisionali

Introduciamo le seguenti $n\,m$ variabili non negative:

$$x_{vt} = \text{potenza assegnata al veicolo } v \text{ nell'ora } t \text{ (kW)},$$
$$\forall v \in \{1, 2, \ldots, n\},\ \forall t \in \{1, 2, \ldots, m\}.$$

Usando queste variabili, un modello LP per il problema è il seguente:

Modello LP di ricarica a costo minimo

$$\min \quad \sum_{v=1}^{n} \sum_{t=1}^{m} \pi_t \, \Delta t \, x_{vt} \tag{5.1a}$$

$$\text{soggetto a} \quad \eta \sum_{t=1}^{m} \Delta t \, x_{vt} \geq e_v, \qquad \forall v \in \{1, 2, \ldots, n\}, \tag{5.1b}$$

$$\sum_{v=1}^{n} x_{vt} + b_t \leq k, \qquad \forall t \in \{1, 2, \ldots, m\}, \tag{5.1c}$$

$$x_{vt} \leq a_{vt} \, \bar{p}_v, \qquad \forall v \in \{1, 2, \ldots, n\},\ \forall t \in \{1, 2, \ldots, m\}, \tag{5.1d}$$

$$x_{vt} \geq 0, \qquad \forall v \in \{1, 2, \ldots, n\},\ \forall t \in \{1, 2, \ldots, m\}. \tag{5.1e}$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (5.1a) minimizza la spesa energetica della notte: la potenza x_{vt} erogata per la durata Δt costa $\pi_t \, \Delta t \, x_{vt}$ euro;
- i vincoli lineari (5.1b) assicurano che ogni veicolo riceva tutta l'energia richiesta entro la partenza; il rendimento η sta a sinistra: dell'energia prelevata solo la frazione η finisce in batteria (n vincoli lineari);
- i vincoli lineari (5.1c) impongono che il prelievo complessivo (ricarica più carico di base) non superi mai la potenza del contatore: sono i vincoli che accoppiano i veicoli tra loro (m vincoli lineari);
- i vincoli lineari (5.1d) limitano la potenza di ogni veicolo al suo caricatore e la azzerano nelle ore in cui il veicolo è assente ($a_{vt} = 0$): la disponibilità è un *dato*, non una decisione, ed è per questo che non servono variabili binarie ($n\,m$ vincoli lineari);
- i vincoli (5.1e) definiscono le variabili del modello.

Varianti: peak shaving ($\min z$ con $\sum_{v=1}^{n} x_{vt} + b_t \le z$ per ogni $t \in \{1, 2, \dots, m\}$); multiobiettivo costo $+\,\rho\, z$.

Spiegazione riga per riga

Vincolo di energia scritto come $\ge$. Caricare più del necessario è ammesso ma mai conveniente (il prezzo è positivo), quindi all'ottimo vale con uguaglianza.

5.3 Esempio svolto a mano

Un veicolo, due ore

Servono $e = 10$ kWh ($\eta = 1$); prezzi $\pi_1 = 0{,}10$ e $\pi_2 = 0{,}20$ €/kWh; potenza massima $\bar{p} = 8$ kW.
Caso $\bar{p} = 8$. Nell'ora economica si carica al massimo (8 kWh); i restanti 2 kWh vanno nell'ora cara. Costo: $8 \cdot 0{,}10 + 2 \cdot 0{,}20 = 1{,}20$ €. *Prezzo ombra del fabbisogno*: l'undicesimo kWh finirebbe nell'ora cara $\Rightarrow$ 0,20 €/kWh. *Prezzo ombra del limite di potenza* nell'ora 1: un kW in più permetterebbe di spostare un kWh dall'ora cara a quella economica $\Rightarrow$ 0,10 €/kW.
Caso $\bar{p} = 12$. Tutto nell'ora economica: costo 1,00 €, duale del fabbisogno 0,10, duale della potenza 0 (vincolo non attivo). Il pattern "il duale è il prezzo dell'ora marginale" è la chiave di lettura di tutto il capitolo.

5.4 Caso di studio

Sei furgoni con finestre di ricarica notturne diverse (dati in `dati/ev_flotta.csv`), prezzi orari stilizzati con picco serale, carico di base dell'edificio, contatore da $k = 120$ kW, $\eta = 0{,}95$.

```
x = m.addVars(veicoli, ore, name="x")
for v in veicoli:
    for t in ore:
        x[v, t].UB = P[v] * disp[v, t]     # 0 se non collegato
v_ene  = m.addConstrs((eta * gp.quicksum(x[v, t] for t in ore)
                       >= E[v] for v in veicoli), name="energia")
```

```
v_rete = m.addConstrs((gp.quicksum(x[v, t] for v in veicoli)
                       + base[t] <= C for t in ore), name="rete")
m.setObjective(gp.quicksum(prezzo[t] * x[v, t]
                           for v in veicoli for t in ore), GRB.MINIMIZE
                              )
```

5.5 Risultati

```
Costo minimo    : 20,36 EUR/notte   picco di prelievo 103,4 kW (limite 120)
Peak shaving    : picco minimo possibile 68,0 kW
Prezzi ombra del fabbisogno (EUR/kWh):
  V1 0,0947   V2 0,0842   V3 0,0842   V4 0,0947   V5 0,0947   V6 0,0842
```

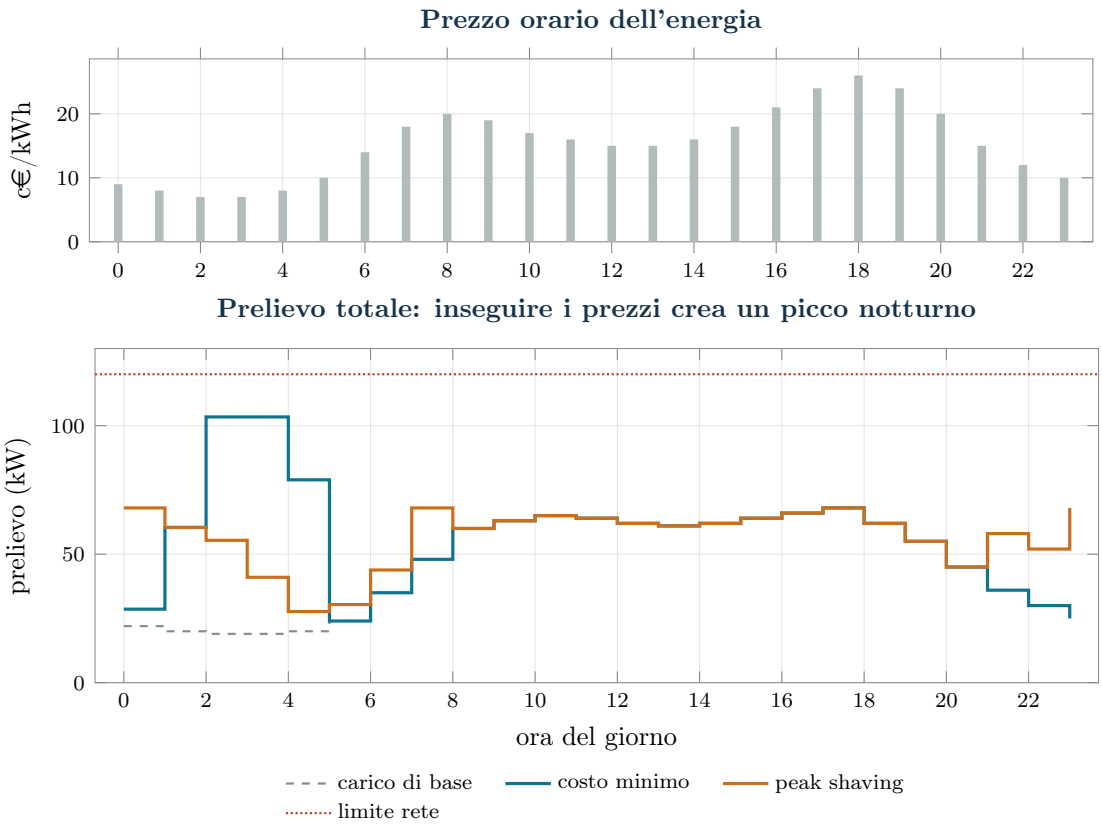


Figura 5.1: Sopra: prezzo orario dell'energia (barre grigie), basso di notte e con picco serale. Sotto: prelievo totale (base + ricarica) ora per ora nelle due strategie: il costo minimo (teal) concentra la ricarica nelle ore economiche e crea un picco notturno di 103 kW; il peak shaving (arancione) distribuisce la ricarica e tiene il prelievo a 68 kW, sotto il limite di rete (linea rossa).

I duali raccontano le ore marginali

I prezzi ombra del fabbisogno valgono $0{,}0842 = 0{,}08/\eta$ oppure $0{,}0947 = 0{,}09/\eta$: sono i prezzi delle *ore marginali* di ciascun veicolo (l'ora più economica ancora libera nella sua finestra), corretti per il rendimento. V2, V3 e V6 hanno finestre che coprono le ore a 0,08; V1, V4 e V5 arrivano tardi o partono presto e la loro ora marginale costa 0,09. Un kWh di fabbisogno in più costa quindi 8,4–9,5 centesimi a seconda del veicolo.

5.6 Analisi di sensitività

```
Compromesso costo + rho*picco:
  rho = 0,00: costo 20,36  picco 103,4     rho = 0,20: costo 22,15  picco 68,0
  rho = 0,05: costo 20,69  picco  86,5     rho = 0,50: costo 22,15  picco 68,0
  rho = 0,10: costo 21,39  picco  74,9
Capacita' del contatore:
  k = 60, 65 kW: INAMMISSIBILE     k = 80: 21,09     k = 120: 20,36
  k = 70 kW    : 21,93             k = 90: 20,63
```

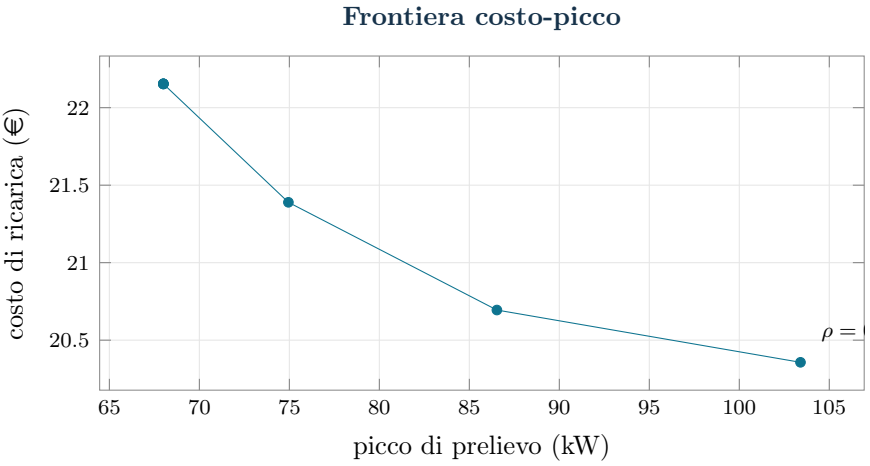


Figura 5.2: Frontiera costo-picco ottenuta variando il peso ρ del picco nell'obiettivo: ogni punto è una strategia di ricarica ottima. Ridurre il picco da 103 a 75 kW costa circa un euro a notte; il punto $\rho \geq 0{,}2$ raggiunge il picco minimo possibile (68 kW) al costo più basso che lo consente.

Insight manageriale

Tagliare il picco da 103 a 75 kW (−28%) costa solo un euro a notte; arrivare a 68 kW ne costa meno di due. Se la tariffa di potenza o l'upgrade del contatore costano più di così, il peak shaving si ripaga da solo. Nota: il minimax puro (picco 68) senza termine di costo spenderebbe 27,79 € — il compromesso con $\rho = 0{,}2$ ottiene *lo stesso picco* a 22,15 €: mai ottimizzare un solo obiettivo quando ce ne sono due.

Nel vincolo `quicksum(x) + base[t] <= C`, Gurobi sposta la costante `base[t]` nel termine noto: il RHS memorizzato è `C - base[t]`. La prima versione della nostra analisi di sensitività scriveva `v.RHS = nuovaC` e otteneva costi identici per ogni capacità — stava in realtà allentando il vincolo di `base[t]`. La forma corretta è `v.RHS = nuovaC - base[t]`. Quando una sensitività non cambia nulla, sospettare prima del proprio codice che del modello.

Un profilo di prelievo *regolare*, che penalizza le variazioni del prelievo da un'ora all'altra, richiede un termine quadratico nell'obiettivo e non è più un LP: è la variante quadratica di questo modello, nel modulo *Modellazione PNL*, pagina «Ricarica di veicoli elettrici con profilo regolare».

5.7 In sintesi

In sintesi

- La ricarica intelligente è un LP puro: si decide quanta potenza, e la disponibilità è un dato, non una variabile binaria.
- Le ore economiche si riempiono per prime; il prezzo ombra del fabbisogno di energia è il prezzo dell'ora marginale diviso per η.
- Costo $+\,\rho$ picco traccia la frontiera costo-picco: il picco scende molto per un piccolo aumento di costo.
- Le costanti dei vincoli finiscono nel termine noto: se un'analisi di sensitività non cambia nulla, si controlla prima il codice.

5.8 Esercizi

Esercizio 5.1 — verifica

Verificare per perturbazione il prezzo ombra del fabbisogno di V1 (portare e_1 da 46 a 47 kWh) e spiegare perché vale 0,09/0,95.

Esercizio 5.2 — arrivo in ritardo

V3 arriva alle 23 invece che alle 20: come cambiano costo e picco? Quale prezzo ombra aumenta?

Esercizio 5.3 — emissioni

Con intensità carbonica oraria g_t (alta la sera, bassa di notte), tracciare la frontiera costo-emissioni. Le ore economiche sono anche le più pulite in questo caso?

Esercizio 5.4 — capacità minima

Trovare per bisezione la capacità minima k che mantiene il problema ammissibile, e spiegare quale combinazione di finestre e $\bar{p}_v$ la determina.

Capitolo 6

Data Envelopment Analysis

Classe: LP **Script:** `python/lab06_dea.py`
Pagina online

Prerequisiti

- Dal modulo *Programmazione lineare*, Capitolo 9 «Teoria della dualità»: il duale di un LP e la dualità forte, da cui viene la forma di inviluppo.
- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: i prezzi ombra e la loro verifica per perturbazione.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- definire l'efficienza relativa come rapporto pesato fra output e input;
- trasformare il modello frazionario in un LP e scriverne il duale;
- leggere nel duale le unità di riferimento e gli obiettivi di miglioramento;
- separare l'inefficienza tecnica da quella di scala.

6.1 Motivazione gestionale

Una banca ha dodici filiali. Ognuna usa personale e costi operativi e produce prestiti e depositi. La direzione vuole sapere quali filiali lavorano bene e quali no, e di quanto dovrebbero ridurre le risorse quelle che lavorano peggio.

Il confronto non è banale: gli output non hanno un prezzo comune, e una filiale grande non è inefficiente solo perché spende di più. Serve un indice che confronti ogni filiale con le *migliori pratiche osservate*, senza fissare a priori quanto vale un milione di depositi rispetto a un milione di prestiti.

La *Data Envelopment Analysis* (DEA) costruisce questo indice risolvendo un LP per ogni filiale. Si usa per confrontare ospedali, scuole, uffici pubblici, impianti, reparti: ovunque ci siano unità simili con più input e più output.

6.2 Unità, input, output ed efficienza

Le unità confrontate si chiamano DMU (*Decision Making Unit*, unità decisionale). Ognuna trasforma gli stessi tipi di input negli stessi tipi di output. I dati sono i seguenti.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di DMU; le DMU sono indicizzate da $j \in \{1, 2, \dots, n\}$
m	$\in \mathbb{Z}_{\geq 1}$	numero di input; gli input sono indicizzati da $i \in \{1, 2, \dots, m\}$
p	$\in \mathbb{Z}_{\geq 1}$	numero di output; gli output sono indicizzati da $r \in \{1, 2, \dots, p\}$
x_{ij}	$\in \mathbb{Q}_{>0}$	quantità dell'input i usata dalla DMU j
y_{rj}	$\in \mathbb{Q}_{>0}$	quantità dell'output r prodotta dalla DMU j
o	$\in \{1, 2, \dots, n\}$	la DMU che si sta valutando

Qui x e y sono *dati*, non variabili: è la notazione standard della DEA, dichiarata una volta per tutte. Gli input e gli output si sommano con dei pesi non negativi: v_i per l'input i e u_r per l'output r.

Definizione 6.1 — Efficienza relativa

Dati i pesi $u_r \geq 0$ e $v_i \geq 0$, non tutti nulli, il *rapporto di efficienza* della DMU j è

$$E_j(u, v) = \frac{\sum_{r=1}^{p} u_r \, y_{rj}}{\sum_{i=1}^{m} v_i \, x_{ij}}.$$

L'*efficienza relativa* della DMU o è il massimo di $E_o(u, v)$ sui pesi per cui nessuna DMU ha rapporto maggiore di 1. La DMU o è *efficiente* se la sua efficienza relativa vale 1.

Ogni DMU sceglie quindi i pesi che la mettono nella luce migliore. Se anche con i pesi più favorevoli resta sotto 1, c'è almeno un'altra DMU (o una combinazione di DMU) che fa meglio con quei pesi: l'inefficienza non dipende da una scelta arbitraria dei prezzi.

6.3 Il modello CCR orientato all'input

Il modello prende il nome da Charnes, Cooper e Rhodes (CCR), che lo hanno proposto. È orientato all'input: misura di quanto si possono ridurre gli input lasciando invariati gli output.

Il problema a parole. *Decidiamo* i pesi degli input e degli output. *L'obiettivo* è massimizzare il rapporto di efficienza della DMU o. *I vincoli* impongono che, con quegli stessi pesi, nessuna DMU abbia rapporto maggiore di 1.

Variabili decisionali

Introduciamo le seguenti $p + m$ variabili non negative:

$$\begin{cases} u_r = \text{peso dell'output } r, & \forall r \in \{1, 2, \dots, p\}, \\ v_i = \text{peso dell'input } i, & \forall i \in \{1, 2, \dots, m\}. \end{cases}$$

Usando queste variabili, il modello frazionario per la DMU o è il seguente:

Modello CCR frazionario per la DMU o

$$\max \quad \frac{\sum_{r=1}^{p} u_r\, y_{ro}}{\sum_{i=1}^{m} v_i\, x_{io}} \tag{6.1a}$$

$$\text{soggetto a} \quad \frac{\sum_{r=1}^{p} u_r\, y_{rj}}{\sum_{i=1}^{m} v_i\, x_{ij}} \le 1, \qquad \forall j \in \{1, 2, \dots, n\}, \tag{6.1b}$$

$$u_r \ge 0, \qquad \forall r \in \{1, 2, \dots, p\}, \tag{6.1c}$$

$$v_i \ge 0, \qquad \forall i \in \{1, 2, \dots, m\}. \tag{6.1d}$$

Descrizione della funzione obiettivo e dei vincoli del modello frazionario:

- la funzione obiettivo (6.1a) massimizza il rapporto di efficienza della DMU o;
- i vincoli (6.1b) impongono che con gli stessi pesi nessuna DMU, o compresa, superi il rapporto 1 (n vincoli frazionari);
- i vincoli (6.1c) e (6.1d) definiscono le variabili del modello.

La trasformazione di Charnes–Cooper

Il modello (6.1) non è lineare, ma ha una proprietà decisiva: se si moltiplicano tutti i pesi per lo stesso $t > 0$, nessun rapporto cambia. Si può quindi scegliere t in modo che il denominatore dell'obiettivo valga 1.

Con $\sum_{i=1}^{m} v_i\, x_{io} = 1$ l'obiettivo diventa il solo numeratore, e ogni vincolo (6.1b), moltiplicato per il suo denominatore positivo, diventa lineare. Si ottiene l'LP seguente, detto *forma dei moltiplicatori*.

Modello CCR, forma dei moltiplicatori, per la DMU o

$$\max \quad \sum_{r=1}^{p} u_r\, y_{ro} \tag{6.2a}$$

$$\text{soggetto a} \quad \sum_{i=1}^{m} v_i\, x_{io} = 1, \tag{6.2b}$$

$$\sum_{r=1}^{p} u_r\, y_{rj} - \sum_{i=1}^{m} v_i\, x_{ij} \le 0, \qquad \forall j \in \{1, 2, \dots, n\}, \tag{6.2c}$$

$$u_r \ge 0, \qquad \forall r \in \{1, 2, \dots, p\}, \tag{6.2d}$$

$$v_i \ge 0, \qquad \forall i \in \{1, 2, \dots, m\}. \tag{6.2e}$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (6.2a) massimizza l'output pesato della DMU o, che è il suo rapporto di efficienza perché l'input pesato vale 1;
- il vincolo lineare (6.2b) normalizza a 1 l'input pesato della DMU o (1 vincolo lineare);

- i vincoli lineari (6.2c) impongono che per ogni DMU l'output pesato non superi l'input pesato, cioè che il rapporto non superi 1 (n vincoli lineari);
- i vincoli (6.2d) e (6.2e) definiscono le variabili del modello.

Il valore ottimo $\tilde{h}_o$ dell'LP (6.2) è l'efficienza relativa della DMU o, e $0 < \tilde{h}_o \leq 1$. Per valutare n DMU si risolvono n LP, uno per DMU: cambiano solo l'obiettivo e il vincolo di normalizzazione.

6.4 La forma di inviluppo: il duale

L'LP (6.2) è un problema di massimo con un vincolo di uguaglianza e n vincoli $\leq$. Il suo duale ha una variabile libera per l'uguaglianza e una variabile non negativa per ogni vincolo $\leq$ (capitolo 2; la teoria completa è nel capitolo Teoria della dualità del corso di Programmazione lineare).

Per aderire alla letteratura della DEA le variabili duali non si chiamano π: introduciamo le seguenti $1 + n$ variabili duali,

$$\begin{cases} \theta = \text{variabile duale del vincolo (6.2b)}, \\ \lambda_j = \text{variabile duale del vincolo (6.2c) della DMU } j, \quad \forall j \in \{1, 2, \ldots, n\}. \end{cases}$$

Il duale si scrive con le regole di conversione e si chiama *forma di inviluppo*.

Modello CCR, forma di inviluppo, per la DMU o

$$\min \quad \theta \tag{6.3a}$$

$$\text{soggetto a} \quad \theta\, x_{io} - \sum_{j=1}^{n} \lambda_j\, x_{ij} \geq 0, \qquad \forall i \in \{1, 2, \ldots, m\}, \tag{6.3b}$$

$$\sum_{j=1}^{n} \lambda_j\, y_{rj} \geq y_{ro}, \qquad \forall r \in \{1, 2, \ldots, p\}, \tag{6.3c}$$

$$\theta \gtreqless 0, \tag{6.3d}$$

$$\lambda_j \geq 0, \qquad \forall j \in \{1, 2, \ldots, n\}. \tag{6.3e}$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (6.3a) minimizza il fattore θ con cui si contraggono tutti gli input della DMU o;
- i vincoli lineari (6.3b) impongono che la combinazione di DMU pesata con i λ_j usi, per ogni input, al più θ volte l'input della DMU o (m vincoli lineari; vengono dalle variabili v_i);
- i vincoli lineari (6.3c) impongono che la stessa combinazione produca, per ogni output, almeno quanto la DMU o (p vincoli lineari; vengono dalle variabili u_r);
- i vincoli (6.3d) e (6.3e) definiscono le variabili del modello.

Che cosa dice la forma di inviluppo

La combinazione $\sum_{j=1}^{n} \lambda_j \, (x_{\cdot j}, y_{\cdot j})$ è una DMU *virtuale* costruita con le DMU osservate. Il modello cerca la DMU virtuale che produce almeno gli output di o con la frazione più piccola θ dei suoi input.
Con $\lambda_o = 1$ e gli altri $\lambda_j = 0$ si ha $\theta = 1$: quindi $\tilde{\theta} \leq 1$. Per la dualità forte $\tilde{\theta} = \tilde{h}_o$: la contrazione radiale ottima degli input è l'efficienza relativa della definizione.

La soluzione ottima della forma di inviluppo dà tre informazioni gestionali:

- l'*insieme di riferimento* (i *peer*) della DMU o: $\{j \in \{1, 2, \ldots, n\} : \tilde{\lambda}_j > 0\}$, le DMU efficienti a cui o deve ispirarsi;
- i *target* degli input e degli output, cioè la proiezione di o sulla frontiera efficiente, che si scrivono dopo aver introdotto gli scarti;
- i prezzi ombra dei vincoli (6.3b) e (6.3c), che sono i pesi ottimi $\tilde{v}_i$ e $\tilde{u}_r$ della forma dei moltiplicatori (uno dei vettori ottimi, se non è unico).

6.5 Il modello BCC e l'efficienza di scala

Il modello CCR assume rendimenti di scala costanti: se una DMU è efficiente, lo è anche la DMU che usa il doppio degli input e produce il doppio degli output. Per filiali di dimensioni molto diverse l'ipotesi è forte.

Il modello di Banker, Charnes e Cooper (BCC) assume rendimenti di scala variabili: aggiunge alla forma di inviluppo il vincolo che i λ_j sommino a 1. La DMU virtuale diventa una combinazione *convessa* delle DMU osservate, cioè una DMU di dimensione comparabile.

Modello BCC, forma di inviluppo, per la DMU o

$$\min \quad \theta \tag{6.4a}$$

$$\text{soggetto a} \quad \theta \, x_{io} - \sum_{j=1}^{n} \lambda_j \, x_{ij} \geq 0, \qquad \forall i \in \{1, 2, \ldots, m\}, \tag{6.4b}$$

$$\sum_{j=1}^{n} \lambda_j \, y_{rj} \geq y_{ro}, \qquad \forall r \in \{1, 2, \ldots, p\}, \tag{6.4c}$$

$$\sum_{j=1}^{n} \lambda_j = 1, \tag{6.4d}$$

$$\theta \gtreqless 0, \tag{6.4e}$$

$$\lambda_j \geq 0, \qquad \forall j \in \{1, 2, \ldots, n\}. \tag{6.4f}$$

Descrizione dei vincoli che cambiano rispetto al modello CCR:

- la funzione obiettivo (6.4a) e i vincoli (6.4b) e (6.4c) sono quelli del modello CCR ($m + p$ vincoli lineari);
- il vincolo lineare (6.4d) chiede che la DMU virtuale sia una combinazione convessa delle DMU osservate (1 vincolo lineare);

- i vincoli (6.4e) e (6.4f) definiscono le variabili del modello.

Il vincolo (6.4d) restringe la regione ammissibile, e in un problema di minimo una regione più piccola non migliora mai l'ottimo: $\tilde{\theta}^{\text{BCC}} \geq \tilde{\theta}^{\text{CCR}}$. La differenza fra le due misure è dovuta alla dimensione.

Definizione 6.2 — Efficienza di scala

L'*efficienza di scala* della DMU o è

$$\text{SE}_o = \frac{\tilde{\theta}_o^{\text{CCR}}}{\tilde{\theta}_o^{\text{BCC}}} \in (0, 1].$$

Vale 1 quando la DMU opera alla scala più produttiva; è minore di 1 quando una parte dell'inefficienza dipende solo dalla dimensione della DMU.

Il verso dell'inefficienza di scala si legge nella forma di inviluppo CCR: se $\sum_{j=1}^{n} \tilde{\lambda}_j < 1$ la DMU è troppo piccola (rendimenti crescenti), se $\sum_{j=1}^{n} \tilde{\lambda}_j > 1$ è troppo grande (rendimenti decrescenti). La lettura vale quando la soluzione ottima è unica.

6.6 Scarti ed efficienza debole e forte

La contrazione radiale riduce tutti gli input della stessa frazione. Dopo la contrazione può restare un margine su un singolo input (o un output prodotto in più dalla DMU virtuale): sono gli *scarti*. Si calcolano con una seconda fase, a $\tilde{\theta}$ fissato.

Introduciamo le seguenti $n + m + p$ variabili non negative:

$$\begin{cases} \lambda_j = \text{peso della DMU } j \text{ nella DMU virtuale}, & \forall j \in \{1, 2, \ldots, n\}, \\ s_i^- = \text{scarto dell'input } i, & \forall i \in \{1, 2, \ldots, m\}, \\ s_r^+ = \text{scarto dell'output } r, & \forall r \in \{1, 2, \ldots, p\}. \end{cases}$$

Seconda fase: massimi scarti per la DMU o

$$\max \quad \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{p} s_r^+ \tag{6.5a}$$

$$\text{soggetto a} \quad \sum_{j=1}^{n} \lambda_j \, x_{ij} + s_i^- = \tilde{\theta} \, x_{io}, \qquad \forall i \in \{1, 2, \ldots, m\}, \tag{6.5b}$$

$$\sum_{j=1}^{n} \lambda_j \, y_{rj} - s_r^+ = y_{ro}, \qquad \forall r \in \{1, 2, \ldots, p\}, \tag{6.5c}$$

$$\lambda_j \geq 0, \qquad \forall j \in \{1, 2, \ldots, n\}, \tag{6.5d}$$

$$s_i^- \geq 0, \qquad \forall i \in \{1, 2, \ldots, m\}, \tag{6.5e}$$

$$s_r^+ \geq 0, \qquad \forall r \in \{1, 2, \ldots, p\}. \tag{6.5f}$$

Descrizione della funzione obiettivo e dei vincoli:

- la funzione obiettivo lineare (6.5a) massimizza la somma degli scarti;

- i vincoli lineari (6.5b) scrivono come uguaglianze i vincoli sugli input della forma di inviluppo, con $\theta = \tilde{\theta}$ (m vincoli lineari);
- i vincoli lineari (6.5c) fanno lo stesso per gli output (p vincoli lineari);
- i vincoli (6.5d), (6.5e) e (6.5f) definiscono le variabili del modello.

Con la soluzione della seconda fase i target della DMU o sono

$$\hat{x}_{io} = \tilde{\theta}\, x_{io} - \tilde{s}_i^-, \quad \forall i \in \{1, 2, \dots, m\}, \qquad \hat{y}_{ro} = y_{ro} + \tilde{s}_r^+, \quad \forall r \in \{1, 2, \dots, p\}.$$

Per il modello BCC si aggiunge il vincolo $\sum_{j=1}^{n} \lambda_j = 1$ e si usa il $\tilde{\theta}$ del modello BCC.

Definizione 6.3 — Efficienza debole e forte

La DMU o è *fortemente efficiente* se $\tilde{\theta} = 1$ e tutti gli scarti ottimi della seconda fase sono nulli. È *debolmente efficiente* se $\tilde{\theta} = 1$ ma almeno uno scarto è positivo.

Una DMU debolmente efficiente non può ridurre tutti gli input nella stessa proporzione, ma può ridurne almeno uno senza perdere output: un'altra DMU fa lo stesso con meno.

6.7 Esempio svolto a mano

Cinque unità, un input e un output

Cinque DMU usano un solo input x e producono un solo output y:

DMU	1	2	3	4	5
x_{1j}	2	4	6	5	2
y_{1j}	2	6	7	3	1

La forma dei moltiplicatori (6.2) per la DMU 1, a coefficienti staccati, è:

$$\begin{aligned} \max \quad & 2\,u_1 \\ \text{soggetto a} \quad & 2\,v_1 = 1 \\ & 2\,u_1 - 2\,v_1 \le 0 \\ & 6\,u_1 - 4\,v_1 \le 0 \\ & 7\,u_1 - 6\,v_1 \le 0 \\ & 3\,u_1 - 5\,v_1 \le 0 \\ & u_1 - 2\,v_1 \le 0 \\ & u_1,\ v_1 \ge 0 \end{aligned}$$

- **Passo 1 — i valori del solver.** Risolvendo i cinque LP CCR e i cinque LP BCC si ottiene:

  $$\tilde{\theta}^{\text{CCR}} = (0{,}6667;\ 1;\ 0{,}7778;\ 0{,}4;\ 0{,}3333), \qquad \tilde{\theta}^{\text{BCC}} = (1;\ 1;\ 1;\ 0{,}5;\ 1).$$

  Per la DMU 1 i pesi ottimi sono $\tilde{u}_1 = 0{,}3333$ e $\tilde{v}_1 = 0{,}5$, e nella forma di inviluppo l'unico peer è la DMU 2 con $\tilde{\lambda}_2 = 0{,}3333$. Per la DMU 5 il modello BCC dà uno scarto di output $\tilde{s}_1^+ = 1$.

- **Passo 2 — verifica del CCR.** Con un solo input e un solo output i pesi si semplificano nel rapporto, e l'efficienza è il rapporto y/x diviso per il migliore, $6/4 = 1{,}5$

della DMU 2:

$$\tilde{h}_1 = \frac{2/2}{1{,}5} = 0{,}6667, \qquad \tilde{h}_3 = \frac{7/6}{1{,}5} = 0{,}7778, \qquad \tilde{h}_4 = \frac{3/5}{1{,}5} = 0{,}4.$$

- **Passo 3 — verifica della dualità per la DMU 1.** I pesi soddisfano la normalizzazione e danno l'efficienza; il vincolo della DMU 2 è attivo:

$$2 \cdot 0{,}5 = 1, \qquad 2 \cdot 0{,}3333 = 0{,}6667, \qquad 6 \cdot 0{,}3333 - 4 \cdot 0{,}5 = 0.$$

  Nella forma di inviluppo la DMU virtuale $\tilde{\lambda}_2 \cdot (4, 6)$ usa esattamente $\tilde{\theta} \cdot 2$ di input e produce esattamente l'output della DMU 1:

$$0{,}3333 \cdot 4 = 1{,}3333 = 0{,}6667 \cdot 2, \qquad 0{,}3333 \cdot 6 = 2.$$

- **Passo 4 — verifica del BCC per la DMU 4.** Il solver combina le DMU 1 e 2 con $\tilde{\lambda}_1 = 0{,}75$ e $\tilde{\lambda}_2 = 0{,}25$, che sommano a 1:

$$0{,}75 \cdot 2 + 0{,}25 \cdot 4 = 2{,}5 = 0{,}5 \cdot 5, \qquad 0{,}75 \cdot 2 + 0{,}25 \cdot 6 = 3.$$

  L'efficienza di scala della DMU 4 è $0{,}4/0{,}5 = 0{,}8$; quella della DMU 1 è 0,6667 e quella della DMU 3 è 0,7778.

- **Passo 5 — efficienza debole.** La DMU 5 ha $\tilde{\theta}^{\text{BCC}} = 1$: nessuna combinazione convessa usa meno di 2 unità di input. Ma la DMU 1 usa lo stesso input e produce 1 unità in più: lo scarto $\tilde{s}_1^+ = 1$ rende la DMU 5 debolmente efficiente.

6.8 Caso di studio

Dodici filiali di una banca ($n = 12$), con due input ($m = 2$: personale in FTE, *Full-Time Equivalent*, e costi operativi in migliaia di euro) e due output ($p = 2$: prestiti e depositi in milioni di euro). I dati sono sintetici ma realistici: li genera lo script con un seme fisso, da una funzione di produzione a rendimenti di scala decrescenti con un'inefficienza casuale per filiale (file `dati/dea_filiali.csv`).

filiale	1	2	3	4	5	6	7	8	9	10	11	12
personale (FTE)	10	25	23	18	17	27	9	24	12	10	20	30
costi (k€)	720	1420	1130	940	990	1620	640	1300	690	500	1020	1370
prestiti (M€)	19,6	60,7	43,5	31,7	39,6	39,7	21,9	34,5	29,2	23,7	42,8	63,5
depositi (M€)	18,5	58,7	63,0	50,5	43,2	73,1	38,0	72,7	28,7	26,3	45,7	62,9

6.9 Implementazione

Il cuore dello script (versione integrale in `lab06_dea.py`) è la funzione che costruisce la forma di inviluppo; un argomento aggiunge il vincolo di convessità del BCC.

```
def inviluppo(X, Y, o, vrs=False):
    m_, n_ = X.shape           # input x DMU
    p_ = Y.shape[0]            # output
    md = gp.Model("dea_inviluppo")
```

```
    theta = md.addVar(lb=-GRB.INFINITY, name="theta")
    lam = md.addVars(n_, name="lambda")
    c_in = md.addConstrs(
        (theta * X[i, o] - gp.quicksum(lam[j] * X[i, j]
         for j in range(n_)) >= 0 for i in range(m_)), name="input")
    c_out = md.addConstrs(
        (gp.quicksum(lam[j] * Y[r, j] for j in range(n_))
         >= Y[r, o] for r in range(p_)), name="output")
    if vrs:                      # BCC: combinazione convessa
        md.addConstr(lam.sum() == 1, name="convessita")
    md.setObjective(theta, GRB.MINIMIZE)
    md.optimize()
    return md, theta, lam, c_in, c_out
```

La funzione `moltiplicatori` costruisce allo stesso modo l'LP (6.2) e la funzione `scarti` la seconda fase (6.5). Lo script risolve per ogni filiale sia il primale sia il duale e controlla con un `assert` che i due valori ottimi coincidano. Ogni modello ha al più 16 variabili e 13 vincoli.

6.10 Risultati

```
filiale     CCR      BCC    scala  somma lambda (CCR)   rendimenti
      1  0.8055   0.9000   0.8950                0.671   crescenti
      2  1.0000   1.0000   1.0000                1.000   costanti
      3  0.9815   1.0000   0.9815                1.917   decrescenti
      4  0.9233   0.9403   0.9819                1.423   decrescenti
      5  0.9604   0.9611   0.9993                1.450   decrescenti
      6  0.7600   1.0000   0.7600                1.924   decrescenti
      7  1.0000   1.0000   1.0000                1.000   costanti
      8  0.9419   1.0000   0.9419                1.913   decrescenti
      9  1.0000   1.0000   1.0000                1.000   costanti
     10  1.0000   1.0000   1.0000                1.000   costanti
     11  0.8997   0.9470   0.9501                1.641   decrescenti
     12  0.9779   1.0000   0.9779                2.679   decrescenti
Filiali CCR-efficienti: [2, 7, 9, 10]
Filiali BCC-efficienti: [2, 3, 6, 7, 8, 9, 10, 12]
Efficienza CCR media: 0.9375; BCC media: 0.9790
```

Per tutte le dodici filiali il valore ottimo della forma di inviluppo coincide con quello della forma dei moltiplicatori, e il prezzo ombra del vincolo di normalizzazione (6.2b) è uguale a $\tilde{\theta}$: è la dualità forte verificata numericamente.

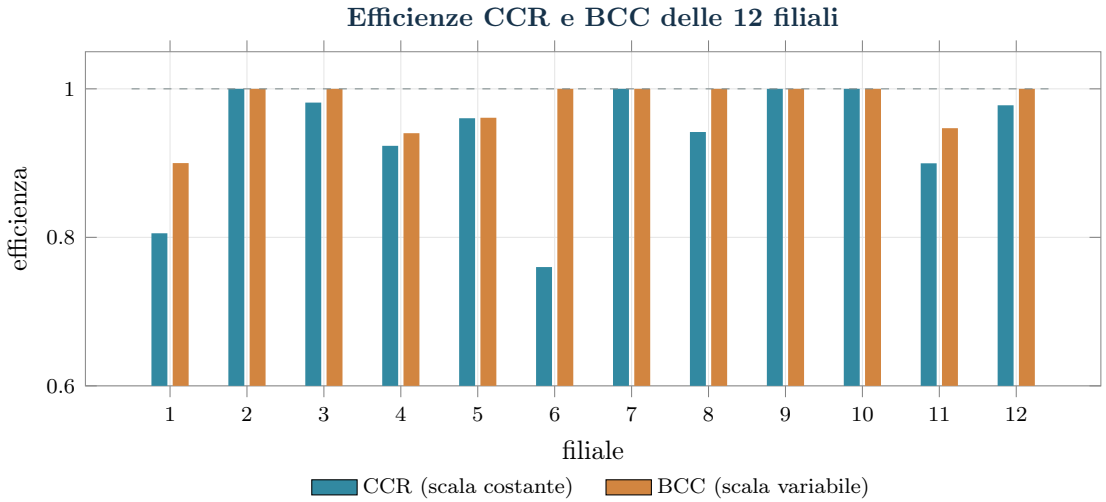


Figura 6.1: Efficienza CCR (teal) ed efficienza BCC (arancio) di ogni filiale; la linea tratteggiata è il valore 1. Le filiali 3, 6, 8 e 12 sono efficienti nel modello BCC ma non nel CCR: la loro inefficienza è tutta di scala.

La tabella seguente riporta, per le filiali inefficienti nel modello CCR, i peer con i loro $\tilde{\lambda}_j$ e i target degli input dopo la seconda fase (file `dati/dea_target.csv`).

filiale	$\tilde{\theta}$	peer (j: $\tilde{\lambda}_j$)	personale	target	costi	target
1	0,8055	9: 0,671	10	8,05	720	463,2
3	0,9815	7: 1,075; 10: 0,842	23	18,10	1130	1109,1
4	0,9233	7: 1,119; 10: 0,304	18	13,11	940	867,9
5	0,9604	7: 0,223; 9: 1,024; 10: 0,203	17	16,33	990	950,8
6	0,7600	7: 1,924	27	17,31	1620	1231,2
8	0,9419	7: 1,913	24	17,22	1300	1224,4
11	0,8997	2: 0,106; 10: 1,535	20	17,99	1020	917,7
12	0,9779	10: 2,679	30	26,79	1370	1339,7

Che cosa fare filiale per filiale

La filiale 6 è la meno efficiente ($\tilde{\theta} = 0{,}76$) ma è efficiente nel modello BCC: non lavora male, è troppo grande. Il suo unico peer è la filiale 7, la più piccola, presa 1,924 volte: la somma dei $\tilde{\lambda}_j$ maggiore di 1 segnala rendimenti decrescenti. La raccomandazione non è "tagliare", ma valutare una divisione in due sportelli.
La filiale 1 è il caso opposto: la somma dei $\tilde{\lambda}_j$ vale 0,671, quindi è troppo piccola, e anche il modello BCC la trova inefficiente (0,90). Dopo la contrazione radiale le resta uno scarto di 116,8 k€ sui costi: il target dei costi è 463,2 k€ contro i 720 attuali.

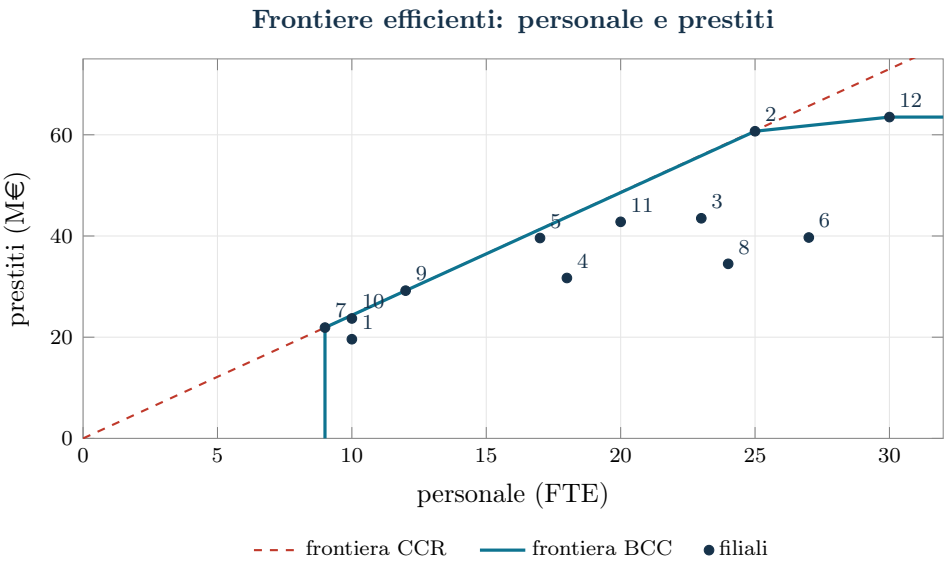


Figura 6.2: Vista a un input e un output (personale e prestiti) delle dodici filiali. La frontiera CCR (tratteggiata) è la semiretta dall'origine con la pendenza migliore, 2,4333 M€ per FTE della filiale 9; la frontiera BCC (continua) passa per le filiali 7, 9, 2 e 12. In questa vista ridotta le efficienze sono diverse da quelle del modello completo.

La figura mostra il significato geometrico dei due modelli. Sotto la semiretta CCR una filiale paga sia l'inefficienza tecnica sia quella di scala; sotto la spezzata BCC paga solo la prima. La filiale 12 sta sulla frontiera BCC ma lontano dalla semiretta: nella vista ha efficienza BCC 1 e CCR 0,870.

6.11 Analisi di sensitività

Prezzi ombra della forma di inviluppo

Per la filiale 6 il solver dà questi prezzi ombra dei vincoli sugli output (6.3c), con i loro intervalli di validità:

```
filiale 6: theta = 0.7600
  output prestiti: y = 39.7 | Pi = 0.00000 | valido per il termine noto in [-
      inf, 42.13]
  output depositi: y = 73.1 | Pi = 0.01040 | valido per il termine noto in
      [68.89, inf]
Verifica: termine noto di depositi da 73.1 a 74.1 -> theta da 0.7600 a 0.7704 (
    variazione +0.01040 = Pi 0.01040)
Confronto: peso u del primale = (0.00000, 0.01040)
```

Il prezzo ombra è il peso del primale

Aumentare il termine noto di un vincolo $\geq$ restringe la regione ammissibile, e in un problema di minimo $\tilde{\theta}$ non può scendere: i prezzi ombra sono ≥ 0. Il prezzo ombra dei depositi, 0,01040, è il peso $\tilde{u}_2$ della forma dei moltiplicatori, come previsto dalla dualità.
La verifica per perturbazione lo conferma: chiedere un milione di depositi in più alza $\tilde{\theta}$ di 0,0104, e il valore resta valido per qualunque termine noto sopra 68,89. Il prezzo ombra dei prestiti è nullo fino a 42,13: la DMU virtuale produce già 2,43 milioni di prestiti più della filiale 6 (lo scarto della seconda fase), e chiedere di più fino a quella soglia non costa nulla.

Stabilità della frontiera

La filiale 7 è peer di cinque filiali inefficienti, come la 10. Togliendola dal confronto e ricalcolando il modello CCR:

```
  filiale  3: CCR 0.9815 -> 1.0000  <- cambia
  filiale  4: CCR 0.9233 -> 0.9795  <- cambia
  filiale  5: CCR 0.9604 -> 0.9791  <- cambia
  filiale  6: CCR 0.7600 -> 0.9156  <- cambia
  filiale  8: CCR 0.9419 -> 1.0000  <- cambia
Senza la filiale 7: 5 filiali cambiano efficienza; nuove efficienti: [3, 8]
```

La DEA misura l'efficienza *relativa* alle DMU osservate: una sola DMU sulla frontiera può cambiare il giudizio su molte altre. Prima di agire sui target conviene controllare i dati delle DMU efficienti più citate come peer (un errore di registrazione le renderebbe "troppo brave") e ripetere l'analisi senza di esse.

6.12 In sintesi

In sintesi

- L'efficienza è il rapporto pesato output/input con i pesi più favorevoli; fissare a 1 l'input pesato dà un LP.
- Il duale, la forma di inviluppo, cerca un'unità virtuale che fa almeno altrettanto con meno: $\tilde{\theta} = \tilde{h}_o$, e le $\tilde{\lambda}_j > 0$ sono i peer.
- Il BCC aggiunge $\sum_{j=1}^{n} \lambda_j = 1$; l'efficienza di scala è $\tilde{\theta}^{\text{CCR}}/\tilde{\theta}^{\text{BCC}}$.
- L'efficienza è relativa alle unità osservate: la frontiera e i peer più citati vanno controllati.

6.13 Esercizi

Esercizio 6.1 — verifica

Con i pesi ottimi della forma dei moltiplicatori per la filiale 4, calcolare il rapporto di efficienza di tutte le dodici filiali. Quali filiali hanno rapporto 1, e che relazione c'è con i peer della filiale 4?

Esercizio 6.2 — orientamento all'output

Scrivere il modello CCR orientato all'output (massimizzare il fattore φ con cui si espandono gli output a input dati). Verificare sui dati che $\tilde{\varphi} = 1/\tilde{\theta}$ per ogni filiale, e che nel modello BCC la relazione in generale non vale.

Esercizio 6.3 — una filiale debolmente efficiente

Aggiungere una filiale 13 uguale alla filiale 7 ma con costi operativi di 700 k€. Calcolarne $\tilde{\theta}$ e gli scarti: è efficiente? In che senso?

Esercizio 6.4 — restrizioni sui pesi

La direzione ritiene che un milione di depositi valga almeno quanto un milione di prestiti. Aggiungere alla forma dei moltiplicatori il vincolo $u_2 \geq u_1$ e ricalcolare le efficienze CCR: quali filiali restano efficienti?

Esercizio 6.5 — super-efficienza

Le quattro filiali CCR-efficienti hanno tutte $\tilde{\theta} = 1$. Per ordinarle, risolvere la forma di inviluppo togliendo la DMU o dall'insieme di riferimento ($\lambda_o = 0$): il valore ottimo può superare 1. Quale filiale è la più solida?

Parte III

Decisioni sotto incertezza

Capitolo 7

Il Newsvendor e le sue varianti

Classe: convesso 1D / LP stocastico a scenari **Script:** `python/lab07_newsvendor.py`
Pagina online

Prerequisiti

- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: i prezzi ombra e i loro intervalli di validità.
- Da questo modulo, Capitolo 3 «Il solver: costruire, risolvere, interpretare»: la costruzione e la soluzione di un LP in `gurobipy`.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- ricavare e applicare la regola del quantile;
- trasformare un problema stocastico in un LP a scenari con il trucco della parte positiva;
- quantificare il valore di modellare l'incertezza (VSS);
- incorporare livelli di servizio e avversione al rischio.

7.1 Motivazione gestionale

Il Newsvendor ("venditore di giornali") descrive una decisione comunissima: scegliere una quantità *prima* di osservare la domanda. Ordinare troppo genera invenduto; troppo poco, vendite perse. Si applica a moda e stagionali, freschi e deperibili, farmaci, capacità alberghiera, componenti di ricambio. È la porta d'ingresso dell'ottimizzazione stocastica, e il suo punto di forza didattico è un risultato netto: **la quantità ottima non è la domanda media**.

Il modello ha più di un secolo e resta il fondamento della gestione delle scorte per prodotti a ciclo singolo. La sua importanza didattica va oltre: è il primo incontro con l'idea che *si decide prima e si scopre poi*, e con gli strumenti per farlo bene (scenari, valore della soluzione stocastica, misure di rischio).

7.2 Il modello base

Il problema a parole. *Decidiamo* una sola quantità q, prima di conoscere la domanda. *L'obiettivo* è minimizzare il costo atteso complessivo dell'errore: quello di aver ordinato troppo (invenduto) più quello di aver ordinato troppo poco (vendite perse). *Vincoli* non ce ne sono nel modello base ($q \geq 0$); compaiono nelle varianti: budget condiviso, livelli di servizio, limiti di capacità.

Dati (input del modello)

Simbolo	Tipo	Significato
D	variabile aleatoria ≥ 0	domanda, con funzione di ripartizione F; eccezione alla convenzione sulle maiuscole, per aderenza alla letteratura: variabile aleatoria maiuscola, realizzazioni minuscole (d_s)
c_u	$\in \mathbb{Q}_{>0}$	costo unitario di *underage* (domanda non servita): margine perso più eventuale penalità (€)
c_o	$\in \mathbb{Q}_{>0}$	costo unitario di *overage* (unità invenduta): costo meno valore di recupero (€)

Con prezzo di vendita p, costo d'acquisto c e valore di recupero $v \leq c$ (tutti in $\mathbb{Q}_{>0}$): $c_u = p - c$ e $c_o = c - v$.

Variabile decisionale

Introduciamo una variabile non negativa:

$$q = \text{quantità ordinata } \textit{prima} \text{ di osservare la domanda } D.$$

Usando questa variabile, il modello per il problema è il seguente:

Modello Newsvendor in forma di costo

$$\min \quad c_o\,\mathbb{E}\big[(q-D)^+\big] + c_u\,\mathbb{E}\big[(D-q)^+\big] \tag{7.1a}$$

$$\text{soggetto a} \quad q \geq 0. \tag{7.1b}$$

Descrizione della funzione obiettivo e del vincolo del modello:

- la funzione obiettivo convessa (7.1a) minimizza il costo atteso dell'errore: eccedenza $(q-D)^+$ pagata c_o e carenza $(D-q)^+$ pagata c_u (il simbolo $(x)^+ = \max\{x, 0\}$ indica la parte positiva);
- il vincolo (7.1b) definisce la variabile del modello.

Se F è continua, l'ottimo soddisfa la **regola del quantile**:

$$F(q^*) = \alpha^* = \frac{c_u}{c_u + c_o} \qquad \Longrightarrow \qquad q^* = F^{-1}(\alpha^*).$$

Da dove viene la regola del quantile

Si ragiona sull'unità marginale: ordinare la q-esima unità in più rende c_u se la domanda la assorbe (probabilità $1 - F(q)$) e costa c_o se resta invenduta (probabilità $F(q)$). Conviene finché $c_u\,(1-F(q)) \geq c_o\,F(q)$, cioè finché $F(q) \leq c_u/(c_u+c_o)$. Se la carenza costa molto più dell'eccedenza, α^* è alto e si ordina un quantile alto; se l'invenduto è costoso (deperibili), si ordina poco.

7.3 Esempio svolto a mano (domanda discreta)

Tre scenari equiprobabili

$D \in \{80, 100, 120\}$ con probabilità 1/3; $c_u = 9$, $c_o = 4$ (quindi $\alpha^* = 9/13 = 0{,}6923$).

- **Passo 1 — costo atteso per i tre candidati naturali.**

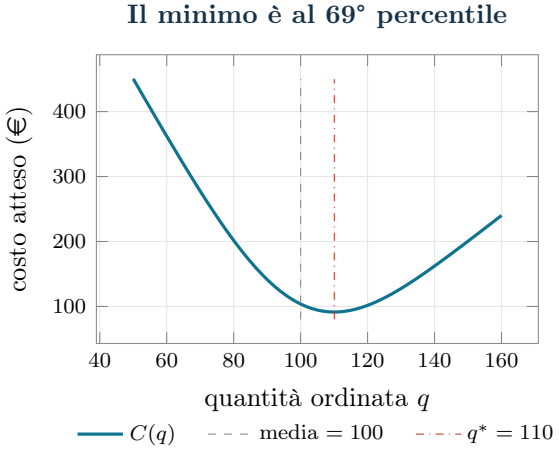
$$C(80) = \tfrac{1}{3}\left[0 + 9\cdot 20 + 9\cdot 40\right] = 180$$
$$C(100) = \tfrac{1}{3}\left[4\cdot 20 + 0 + 9\cdot 20\right] = \tfrac{260}{3} = 86{,}7$$
$$C(120) = \tfrac{1}{3}\left[4\cdot 40 + 4\cdot 20 + 0\right] = 80$$

- **Passo 2 — regola del quantile discreta.** Serve il più piccolo q con $F(q) \geq 0{,}6923$: $F(80) = 1/3$, $F(100) = 2/3 = 0{,}667 < 0{,}6923$, $F(120) = 1$. Quindi $q^* = 120$: coerente con il passo 1.
- **Passo 3 — lettura.** Ordinare il *massimo* storico nonostante la media sia 100: con $c_u = 9 \gg c_o = 4$, restare corti costa più del doppio che restare lunghi. Chi ordina "la media" paga 86,7 contro 80: l'incertezza si gestisce, non si ignora.

7.4 Caso di studio: la regola del quantile in continuo

Panetteria: $p = 15$, $c = 6$, $v = 2$ € $\Rightarrow c_u = 9$, $c_o = 4$; domanda normale con $\mu = 100$, $\sigma = 20$.

```
alpha* = 9/13 = 0,6923
q* = F^-1(0,6923) = 110,05 unita'   (media = 100)
costo atteso in q*: 91,43 EUR    in q = media: 103,72 EUR
```



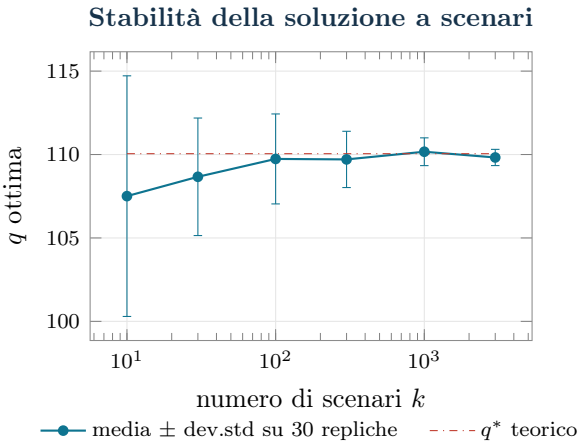


Figura 7.1: A sinistra: il costo atteso $C(q)$ è convesso e il minimo cade in $q^* = 110$ (69° percentile della domanda), non sulla media 100. A destra: la q ottima stimata da k scenari campionati, media e deviazione standard su 30 repliche — con pochi scenari la decisione è instabile.

7.5 La formulazione lineare a scenari

Quando F non è nota in forma chiusa — il caso normale in azienda — si usano gli scenari: dati storici, campioni da una distribuzione stimata, o scenari pesati. Nuovi dati del modello: il numero di scenari $k \in \mathbb{Z}_{\geq 1}$, indicizzati da $s \in \{1, 2, \ldots, k\}$; per ogni scenario, la domanda $d_s \in \mathbb{Q}_{\geq 0}$ e la probabilità $\pi_s \in \mathbb{Q}_{\geq 0}$, con $\sum_{s=1}^{k} \pi_s = 1$.

Oltre a q, introduciamo le seguenti $2\,k$ variabili non negative:

$$\begin{cases} o_s = \text{eccedenza (unità invendute) nello scenario } s \\ u_s = \text{carenza (domanda non servita) nello scenario } s \end{cases} \qquad \forall s \in \{1, 2, \ldots, k\}.$$

Usando queste variabili, un modello LP per il problema è il seguente:

Modello LP del Newsvendor a scenari

$$\min \quad \sum_{s=1}^{k} \pi_s \big(c_o\, o_s + c_u\, u_s\big) \tag{7.2a}$$

$$\text{soggetto a} \quad o_s \geq q - d_s, \qquad \forall s \in \{1, 2, \ldots, k\}, \tag{7.2b}$$

$$u_s \geq d_s - q, \qquad \forall s \in \{1, 2, \ldots, k\}, \tag{7.2c}$$

$$q \geq 0, \tag{7.2d}$$

$$o_s,\ u_s \geq 0, \qquad \forall s \in \{1, 2, \ldots, k\}. \tag{7.2e}$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (7.2a) è la media pesata, sugli scenari, dei costi di eccedenza e di carenza: la versione "a scenari" del costo atteso del modello base;
- i vincoli lineari (7.2b) e (7.2c) definiscono eccedenza e carenza di ogni scenario a partire dall'unica vera decisione q; insieme alle non negatività realizzano le parti positive, come spiegato sotto ($2\,k$ vincoli lineari);
- i vincoli (7.2d) e (7.2e) definiscono le variabili del modello.

Perché è lineare (il trucco delle parti positive)

$(q - d_s)^+$ non è lineare, ma all'*ottimo* la variabile o_s, che compare nell'obiettivo con coefficiente positivo, si schiaccia sul massimo tra $q - d_s$ e 0: il vincolo (7.2b) insieme a $o_s \geq 0$ fa esattamente il lavoro di $\max\{q - d_s, 0\}$ senza variabili binarie. Stesso meccanismo per u_s. Questo trucco — linearizzare le parti positive che l'obiettivo spinge verso il basso — torna identico nel CVaR e nella regressione quantile.

```
q = m.addVar(name="q")
o = m.addVars(S, name="o")      # eccedenza per scenario
u = m.addVars(S, name="u")      # carenza per scenario
```

```
m.addConstrs((o[s] >= q - dom[s] for s in range(S)))
m.addConstrs((u[s] >= dom[s] - q for s in range(S)))
m.setObjective(gp.quicksum((Co*o[s] + Cu*u[s]) / S
                           for s in range(S)), GRB.MINIMIZE)
```

```
LP con k = 600 scenari: q = 108,61
quantile empirico al 69,23%: 108,60   (coincidono: e' un teorema)
VSS: decisione q = E[D] = 100 costa 99,50; q stocastica costa 88,33
     valore della soluzione stocastica = 11,17 EUR per ciclo
```

Insight manageriale

Il *valore della soluzione stocastica* (VSS) risponde alla domanda del capo: "perché complicarsi la vita con gli scenari?". Qui vale 11,17 € a ciclo, l'11% del costo: è il guadagno di *modellare* l'incertezza invece di sostituirla con la media. La figura sulla stabilità dà l'altra metà del messaggio: sotto i 100 scenari la q ottima balla di ± 3 unità tra una stima e l'altra — gli scenari sono anch'essi un campione, e la decisione eredita la loro varianza.

7.6 Livelli di servizio

```
cycle service level 90%: q = 125,63  (costo +23,41 EUR sull'ottimo economico)
cycle service level 95%: q = 132,90  (costo +45,59)
cycle service level 99%: q = 146,53  (costo +95,56)
fill rate in q*: 96,06%   probabilita' di non-stockout in q*: 69,23%
```

In $q^* = 110$ la probabilità di evitare lo stock-out è solo il 69%, ma il *fill rate* (quota di domanda servita in media) è il 96%: misure diverse che i contratti spesso confondono. Promettere "95% di probabilità di copertura totale" costa +45,59 € a ciclo rispetto all'ottimo economico; promettere "95% di fill rate" è quasi gratis. Leggere bene lo SLA prima di firmarlo.

7.7 Avversione al rischio: media-CVaR

Il costo atteso non distingue tra tante piccole perdite e rari disastri. Con la formulazione lineare del CVaR (qui solo anticipata) minimizziamo $(1-\lambda)\,\mathbb{E}[\text{costo}] + \lambda\,\text{CVaR}_{0,90}(\text{costo})$:

```
lambda = 0,00: q = 108,61  costo medio 88,33  CVaR90 245,93
lambda = 0,25: q = 112,45  costo medio 90,37  CVaR90 229,29
lambda = 0,50: q = 114,48  costo medio 92,52  CVaR90 225,49
lambda = 1,00: q = 116,17  costo medio 94,87  CVaR90 224,56
```

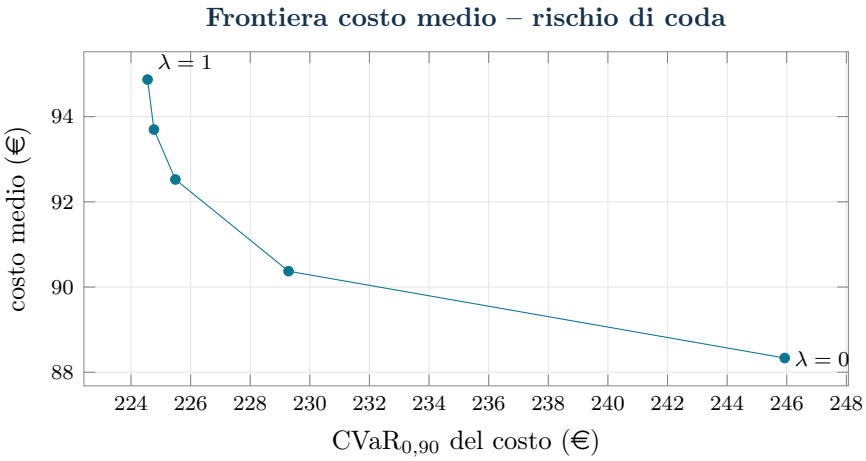


Figura 7.2: Frontiera tra costo medio e $\text{CVaR}_{0,90}$ del costo, ottenuta variando il peso λ dell'avversione al rischio: da $\lambda = 0$ (ottimo in media, coda pesante) a $\lambda = 1$ (protezione massima). Il primo tratto è quasi verticale: molta protezione per pochissimo costo medio.

Il decisore avverso al rischio ordina di più (da 108,6 a 116,2 unità): paga 6,5 € in più in media per tagliare di 21,4 € il costo medio dei peggiori 10% degli scenari. La gran parte della protezione si ottiene già con $\lambda = 0{,}25$: la frontiera è ripida all'inizio e piatta poi.

7.8 Multiprodotto con budget condiviso

Tre dolci con domande *correlate* ($\rho = 0{,}7$: le feste vanno bene o male per tutti insieme) e un budget di produzione di 1200 €:

```
  prodotto   | q senza budget | q con budget
 panettone   |          111,1 |         99,3
 pandoro     |           91,5 |         75,4
 torrone     |           65,9 |         56,8
Spesa: 1200/1200   prezzo ombra del budget: -0,553
```

Spiegazione riga per riga

Il budget vincolante comprime le tre quantità sotto i rispettivi quantili ottimi, ma non proporzionalmente: il taglio dipende dal rapporto c_u/c_o e dal costo unitario di ciascun prodotto. Il duale dice che un euro di budget in più ridurrebbe il costo atteso di 0,55 €: un rendimento del 55% che giustificherebbe quasi qualunque finanziamento — informazione che senza il modello nessuno avrebbe quantificato. La correlazione alta, inoltre, toglie il beneficio di diversificazione: i tre prodotti falliscono insieme, e trattarli come indipendenti sottostimerebbe il rischio.

7.9 In sintesi

In sintesi

- La quantità si sceglie prima di osservare la domanda, e l'ottimo è il quantile $q^* = F^{-1}\big(c_u/(c_u + c_o)\big)$, non la media.
- Con k scenari il newsvendor diventa un LP con $2k + 1$ variabili: le parti positive si

linearizzano senza variabili binarie.

- Il VSS misura il guadagno di modellare l'incertezza; le decisioni stimate da pochi scenari sono instabili.
- Livello di servizio di ciclo e fill rate sono misure diverse; con il mean-CVaR un po' di avversione al rischio dà molta protezione.

7.10 Esercizi

Esercizio 7.1 — verifica

Ricalcolare a mano α^* e q^* se il Comune impone una penalità $b = 3$ € per cliente non servito ($c_u = p - c + b$). Verificare con lo script.

Esercizio 7.2 — dati storici

Usare come scenari le 24 osservazioni di domanda degli ultimi due anni (generarle con lo script) invece della distribuzione: quanto differisce q? Quale approccio consiglierete a un'azienda che ha solo 24 osservazioni?

Esercizio 7.3 — deperibili

Il recupero diventa negativo ($v = -1$ €: smaltire costa): come cambiano α^* e q^*? Tracciare q^* in funzione di $v \in [-2, 4]$.

Esercizio 7.4 — EVPI

Calcolare l'EVPI (*Expected Value of Perfect Information*, valore atteso dell'informazione perfetta): costo medio con q scelta *dopo* aver visto la domanda (cioè $q = d_s$, costo 0) contro costo della soluzione stocastica. Quanto vale al massimo una previsione perfetta della domanda?

Esercizio 7.5 — fill rate vincolato

Aggiungere all'LP il vincolo di fill rate $\sum_{s=1}^{k} \pi_s u_s \le (1 - 0{,}98)\,\mathbb{E}[D]$ e confrontare quantità e costo con il cycle service level al 98%.

Esercizio 7.6 — multiprodotto

Rieseguire il multiprodotto con domande indipendenti ($\rho = 0$) a parità di tutto il resto: il costo atteso ottimo migliora o peggiora? Perché?

Capitolo 8

VaR e CVaR: misurare e ottimizzare il rischio

Classe: LP a scenari **Script:** `python/lab08_var_cvar.py`
Pagina online

Prerequisiti

- Da questo modulo, Capitolo 7 «Il Newsvendor e le sue varianti»: l'LP a scenari, il trucco della parte positiva e l'obiettivo mean-CVaR.
- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: i prezzi ombra e il protocollo di sensitività.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- definire e calcolare VaR e CVaR su una distribuzione discreta;
- usare la formulazione lineare di Rockafellar–Uryasev;
- costruire portafogli e piani logistici protetti dagli scenari peggiori;
- riconoscere i limiti statistici delle stime delle code.

8.1 Motivazione gestionale

Molti modelli ottimizzano un valore medio. Scelta semplice, ma che può nascondere perdite rare e molto elevate. VaR e CVaR descrivono la *coda* della distribuzione delle perdite e trasformano l'avversione al rischio in decisioni quantitative. Il VaR risponde a "qual è una soglia di perdita elevata?"; il CVaR risponde anche a "quanto perdiamo *in media* quando quella soglia viene superata?".

Dopo le crisi finanziarie, "quanto perdiamo negli scenari peggiori?" è diventata una domanda regolamentare oltre che gestionale. VaR e CVaR sono il linguaggio standard con cui banche, assicurazioni e, sempre più, supply chain manager parlano di rischio di coda. La sorpresa didattica è che il CVaR — concetto apparentemente sofisticato — si ottimizza con un semplice LP.

8.2 Definizioni e proprietà

Il problema a parole. Nei modelli di questo capitolo *decidiamo* come prima (quote di portafoglio, quantità, capacità prenotate), ma *l'obiettivo* cambia: non minimizzare la perdita *media*, bensì una misura della sua *coda* — quanto si perde negli scenari peggiori. *I vincoli* restano quelli del problema sottostante, più i vincoli lineari ausiliari che definiscono il CVaR.

Sia L la perdita aleatoria (valori alti = esiti peggiori) e $\alpha \in (0,1)$ un livello di confidenza (tipicamente 0,90–0,99). Come per la domanda del newsvendor, la variabile aleatoria è maiuscola e le sue realizzazioni minuscole (ℓ_s): è l'eccezione, per aderenza alla letteratura, alla nostra convenzione che riserva le maiuscole a insiemi e matrici.

VaR e CVaR

$$\mathrm{VaR}_\alpha(L) = \inf\{\eta \in \mathbb{R} : \mathbb{P}(L \le \eta) \ge \alpha\},$$
$$\mathrm{CVaR}_\alpha(L) = \text{perdita media nella coda peggiore di massa } 1-\alpha.$$

Convessità	il CVaR è convesso nelle decisioni (se L lo è); il VaR in generale no
Subadditività	il CVaR premia la diversificazione; il VaR può violarla
Coda	il VaR non distingue due code con lo stesso quantile ma gravità diversa
Ottimizzazione	il CVaR ammette una formulazione LP a scenari; il VaR no

Dati del modello a scenari: il numero di scenari $k \in \mathbb{Z}_{\ge 1}$, indicizzati da $s \in \{1,2,\dots,k\}$; per ogni scenario, la probabilità $\pi_s \in \mathbb{Q}_{\ge 0}$ (con $\sum_{s=1}^{k} \pi_s = 1$) e la perdita $\ell_s(\boldsymbol{x})$, funzione lineare del vettore delle decisioni $\boldsymbol{x}$, vincolato in un poliedro X; il livello di confidenza $\alpha \in (0,1)$.

Oltre alle decisioni $\boldsymbol{x}$, introduciamo una variabile libera e k variabili non negative:

$$\begin{cases} \eta = \text{soglia candidata di perdita (all'ottimo: un VaR)} \\ \xi_s = \text{eccesso di perdita oltre } \eta \text{ nello scenario } s \end{cases} \qquad \forall s \in \{1,2,\dots,k\}.$$

Usando queste variabili, il modello LP di Rockafellar–Uryasev è il seguente:

Modello LP di Rockafellar–Uryasev per il CVaR

$$\min \quad \eta + \frac{1}{1-\alpha} \sum_{s=1}^{k} \pi_s \, \xi_s \tag{8.1a}$$
$$\text{soggetto a} \quad \xi_s \ge \ell_s(\boldsymbol{x}) - \eta, \qquad \forall s \in \{1,2,\dots,k\}, \tag{8.1b}$$
$$\boldsymbol{x} \in X, \tag{8.1c}$$
$$\eta \gtreqless 0, \tag{8.1d}$$
$$\xi_s \ge 0, \qquad \forall s \in \{1,2,\dots,k\}. \tag{8.1e}$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (8.1a) è la soglia η più la media pesata degli eccessi di coda, amplificata di $1/(1-\alpha)$; *all'ottimo η^* è un VaR e il valore obiettivo è il CVaR*: un solo LP fornisce entrambe le misure;

- i vincoli lineari (8.1b) definiscono gli eccessi: solo gli scenari con perdita oltre la soglia contribuiscono (k vincoli lineari);
- il vincolo (8.1c) raccoglie i vincoli del problema decisionale sottostante (poliedrali);
- i vincoli (8.1d) e (8.1e) definiscono le variabili: η è *libera* (il simbolo $\gtrless$), ξ_s non negative.

Spiegazione riga per riga

La variabile $\xi_s = (\ell_s - \eta)^+$ è l'eccesso di perdita oltre la soglia candidata η (stesso trucco delle parti positive del Newsvendor). Minimizzare su η bilancia due spinte: alzare η riduce gli eccessi ma sposta in su la soglia; il punto di equilibrio è esattamente il quantile α. Se X è un poliedro e ℓ_s è lineare, tutto il modello è un LP.

8.3 Esempio svolto a mano

Sei scenari, $\alpha = 0{,}80$

Perdite equiprobabili $\{2, 4, 5, 7, 12, 20\}$.
VaR. La cumulata raggiunge 0,80 in corrispondenza della quinta perdita ordinata: $\mathrm{VaR}_{0{,}80} = 12$.
CVaR con la formula. La coda peggiore ha massa 0,20; lo scenario da 20 pesa $1/6 = 0{,}167$, e la massa restante 0,033 sta *sul punto* 12:

$$\mathrm{CVaR}_{0{,}80} = 12 + \frac{1}{0{,}20} \cdot \frac{1}{6}\,(20 - 12) = 12 + 6{,}67 = 18{,}67.$$

Verifica con l'LP (script): $\eta^* = 12$, $\xi = (0, 0, 0, 0, 0, 8)$, valore 18,6667. La formula tratta correttamente la massa "a cavallo" del quantile senza selezionare scenari con variabili binarie.
Confronto. Perdita media 8,33; VaR 12; CVaR 18,67: tre numeri, tre storie. La media rassicura, il VaR dà una soglia, il CVaR dice quanto fa male quando va male.

8.4 Caso di studio 1: portafoglio mean-CVaR

Otto titoli azionari settoriali (ENE, FIN, TEC, IND, SAN, CON, UTL, MAT) e $k = 220$ scenari mensili generati con *code grasse* (fattore di mercato t di Student, gradi di libertà 4): il mondo in cui il CVaR mostra il suo valore. Perdita di scenario: $\ell_s(\boldsymbol{x}) = -\sum_{i=1}^{n} r_{is}\, x_i$, dove r_{is} è il rendimento del titolo i nello scenario s. Minimizziamo $\mathrm{CVaR}_{0{,}90}$ con rendimento atteso $\geq 8\%$ annuo.

```
eta = m.addVar(lb=-GRB.INFINITY, name="eta")    # soglia: libera!
xi  = m.addVars(S, name="xi")
m.addConstr(x.sum() == 1)
m.addConstr(gp.quicksum(mu[i]*x[i] for i in range(n)) >= r_min)
m.addConstrs((xi[s] >= -gp.quicksum(R[s, i]*x[i] for i in range(n))
              - eta for s in range(S)))
m.setObjective(eta + gp.quicksum(xi[s] for s in range(S))
                   / (S * (1 - alpha)), GRB.MINIMIZE)
```

```
             | perdita media | VaR90   | CVaR90    (perdite mensili)
  mean-CVaR  |       -0,0067 | 0,0327  | 0,0489
Composizione: ENE 9,9 FIN 1,9 TEC 10,5 SAN 8,6 UTL 46,5 MAT 22,5
```

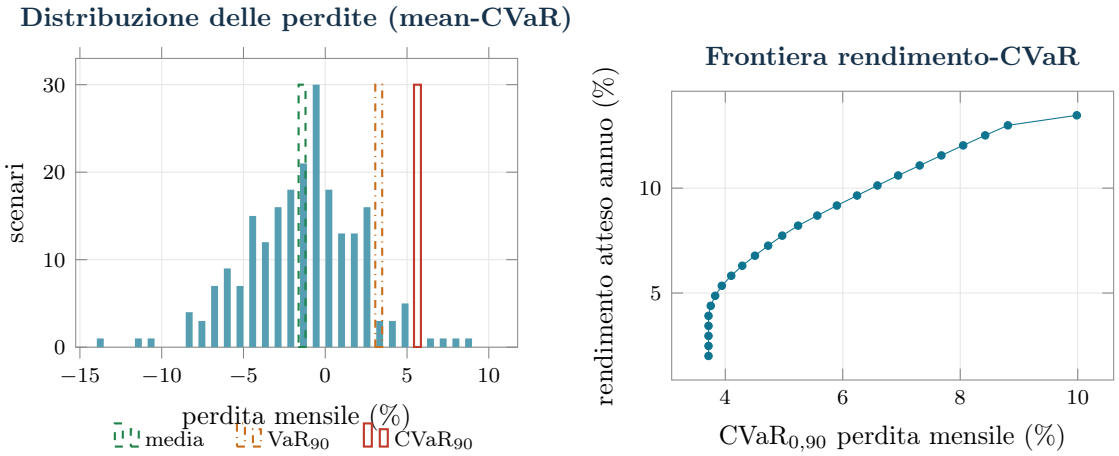


Figura 8.1: A sinistra: istogramma delle perdite mensili del portafoglio mean-CVaR sui 220 scenari, con media (verde), VaR_{90} (arancione) e CVaR_{90} (rosso): il CVaR è la media della coda a destra del VaR. A destra: frontiera rendimento-CVaR ottenuta variando il rendimento minimo richiesto.

Insight manageriale

Con un rendimento atteso dell'8% annuo il portafoglio guadagna in media lo 0,67% al mese, ma nel 10% dei mesi peggiori perde in media il 4,89%: è il numero che la media nasconde. Il CVaR concentra il portafoglio su UTL e MAT e tiene una quota di TEC che, pur mediocre in media, si muove in controtendenza negli scenari peggiori di questo campione. Il CVaR guarda solo dove fa male: le perdite sopra il VaR.

Misurare il rischio con la varianza invece che con il CVaR porta al portafoglio di Markowitz, un modello quadratico: è nel modulo *Modellazione PNL*, pagina «Portafoglio di Markowitz».

8.5 Caso di studio 2: supply chain a due stadi

Prima di osservare lo scenario si *prenota* capacità dai fornitori; poi, visto lo scenario, si decide il ricorso. F1 è economico ma fragile (nel 12% degli scenari la sua disponibilità crolla al 30%); F2 è caro ma affidabile. Penalità di shortage: 40 €/unità; $k = 400$ scenari.

Dati: due fornitori indicizzati da $a \in \{1, 2\}$ (1 economico ma fragile, 2 caro ma affidabile) con costi di prenotazione $c_a^{\text{pre}} \in \mathbb{Q}_{>0}$ e di acquisto $c_a^{\text{uso}} \in \mathbb{Q}_{>0}$; k scenari $s \in \{1, 2, \dots, k\}$ con domanda $d_s \in \mathbb{Q}_{\geq 0}$ e disponibilità $\delta_{as} \in (0, 1]$ (il dato che modella il guasto); penalità di shortage $g \in \mathbb{Q}_{>0}$; peso del rischio $\lambda \in [0, 1]$ e livello $\alpha \in (0, 1)$.

Introduciamo le variabili non negative x_a (capacità prenotata dal fornitore a, *di primo stadio*: uguale in tutti gli scenari), f_{as} (acquisto dal fornitore a nello scenario s) e u_s (domanda non servita nello scenario s), queste ultime *di ricorso*.

Modello a due stadi con CVaR

$$\min \;\; (1-\lambda)\Big[\sum_{a=1}^{2} c_a^{\text{pre}} x_a + \sum_{s=1}^{k} \pi_s\, \phi_s\Big] + \lambda\, \text{CVaR}_\alpha \tag{8.2a}$$

$$\text{soggetto a} \quad f_{as} \le \delta_{as}\, x_a, \quad \forall a \in \{1,2\},\ \forall s \in \{1,2,\dots,k\}, \tag{8.2b}$$

$$f_{1s} + f_{2s} + u_s \ge d_s, \qquad \forall s \in \{1,2,\dots,k\}, \tag{8.2c}$$

$$x_a,\ f_{as},\ u_s \ge 0, \qquad \forall a \in \{1,2\},\ \forall s \in \{1,2,\dots,k\}, \tag{8.2d}$$

dove $\phi_s = \sum_{a=1}^{2} c_a^{\text{uso}} f_{as} + g\, u_s$ è il costo di ricorso dello scenario s e CVaR_α è il CVaR del costo totale, linearizzato con le variabili η, ξ_s di (8.1).

Descrizione della funzione obiettivo e dei vincoli del modello:

- la funzione obiettivo (8.2a) combina, con peso λ, il costo medio (prenotazioni più valore atteso di acquisti e penalità) e il CVaR del costo totale;
- i vincoli lineari (8.2b) impongono che nello scenario s si possa usare il fornitore a solo fino alla capacità prenotata x_a, ridotta dalla disponibilità δ_{as} ($2\,k$ vincoli lineari);
- i vincoli lineari (8.2c) impongono di coprire la domanda con i flussi o, in mancanza, di conteggiarla come shortage u_s, che in obiettivo paga la penalità: il modello non è mai inammissibile, "compra" l'inammissibilità al prezzo della penalità (k vincoli lineari);
- i vincoli (8.2d) definiscono le variabili del modello.

```
lambda = 0,0: prenoto forn.1 161,5  forn.2 134,2
              costo medio 1.724   CVaR90 2.902   servizio 98,7%
lambda = 0,5: prenoto forn.1  92,7  forn.2 212,7
              costo medio 1.803   CVaR90 2.217   servizio 99,9%
lambda = 1,0: prenoto forn.1  68,5  forn.2 236,9
              costo medio 1.906   CVaR90 2.139   servizio 99,0%
```

Insight manageriale

Il decisore neutrale compra dal fornitore economico e accetta che nel 12% degli scenari lo shortage faccia esplodere i costi (CVaR 2.902). Al crescere di λ la capacità migra verso il fornitore affidabile: $+79$ € di costo medio comprano -685 € di CVaR. Questo è il *costo della resilienza*, ed è esattamente il numero che serve in una discussione su rishoring e dual sourcing.

8.6 Analisi di sensitività

I parametri da variare, con l'effetto atteso:

α	più alto ⇒ coda più estrema stimata da meno scenari: servono più scenari per stime stabili
λ	controlla il compromesso prestazione media / protezione
π_s	scenari ripesati o stress test (aggiungere scenari estremi)
$\lvert S\rvert$	stabilità statistica: ripetere con campioni diversi, come nel capitolo 7

Con $\alpha = 0{,}99$ e $k = 220$ scenari la coda contiene solo 2–3 scenari: il CVaR stimato è quasi rumore. Regola pratica: servono almeno qualche decina di scenari *oltre* il quantile. Inoltre con distribuzioni discrete il VaR può non essere unico (η^* è *un* VaR): confrontare sempre η^* con il quantile empirico.

8.7 In sintesi

In sintesi

- Il VaR è una soglia di perdita alta, il quantile α; il CVaR è la perdita media nella coda oltre la soglia.
- Il CVaR è convesso, subadditivo e sensibile alla gravità della coda; il VaR no.
- L'LP di Rockafellar–Uryasev usa una soglia libera η e gli eccessi $\xi_s \geq 0$: all'ottimo η^* è un VaR e il valore è il CVaR.
- Le code estreme stimate da pochi scenari sono rumorose: servono decine di scenari oltre il quantile.

8.8 Esercizi

Esercizio 8.1 — verifica

Ricalcolare a mano VaR e CVaR dell'esempio a 6 scenari con $\alpha = 0{,}90$ e verificare con l'LP. Attenzione: la coda ha massa 0,10 e lo scenario da 20 pesa $0{,}167 > 0{,}10$.

Esercizio 8.2 — subadditività

Costruire un esempio a scenari con due attività in cui $\mathrm{VaR}(L_1 + L_2) > \mathrm{VaR}(L_1) + \mathrm{VaR}(L_2)$ e verificare che il CVaR invece non viola la subadditività.

Esercizio 8.3 — frontiera in α

Per il portafoglio, tracciare il CVaR ottimo per $\alpha \in \{0{,}80;\ 0{,}90;\ 0{,}95;\ 0{,}99\}$ a parità di rendimento: come cambia la composizione? Quando iniziano le instabilità numeriche?

Esercizio 8.4 — CVaR vincolato

Riformulare il portafoglio come $\max \sum_{i=1}^{n} \mu_i x_i$ soggetto a $\mathrm{CVaR}_{0{,}90} \leq \bar{c}$: verificare che variando $\bar{c}$ si percorre la stessa frontiera e leggere il duale del vincolo.

Esercizio 8.5 — stress test

Aggiungere alla supply chain uno scenario estremo (domanda +50% e F1 al 10%) con probabilità 1%: quanto cambia la prenotazione ottima per $\lambda = 0{,}5$? E il CVaR?

Capitolo 9

Arbitraggio e prezzatura senza arbitraggio

Classe: LP **Script:** `python/lab09_arbitraggio.py`

Pagina online

Prerequisiti

- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: la dualità, gli scarti complementari e la dualità forte.
- Da questo modulo, Capitolo 3 «Il solver: costruire, risolvere, interpretare»: le variabili libere in `gurobipy` e la lettura di `Pi`.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- formulare la ricerca di un arbitraggio come LP;
- leggere le probabilità neutrali al rischio nei duali dei vincoli;
- calcolare l'intervallo dei prezzi coerenti di un titolo;
- prezzare un titolo nuovo (un'opzione) in un mercato completo o incompleto.

9.1 Motivazione gestionale

Un *arbitraggio* è una strategia di compravendita che crea denaro dal nulla: incassa oggi senza alcun rischio di perdita domani, oppure non costa nulla oggi e può solo guadagnare domani. In un mercato ben funzionante queste occasioni non possono durare: chiunque le sfrutterebbe su scala illimitata, e i prezzi si aggiusterebbero subito. "Assenza di arbitraggio" è quindi l'ipotesi di lavoro di tutta la finanza dei prezzi: dei desk di trading che cercano incoerenze tra strumenti quotati, dei risk manager che validano i prezzi di listino, di chi deve dare un prezzo a un titolo nuovo coerente con quelli esistenti.

La sorpresa didattica è che tutto questo è *programmazione lineare*: rilevare un arbitraggio è risolvere un LP, e la dualità del capitolo 2 — prezzi ombra, scarti complementari, dualità forte — diventa la teoria della prezzatura. È l'applicazione più pulita della coppia primale-duale di tutta la dispensa.

9.2 Costruzione guidata del modello

Il problema a parole. Il mercato ha due istanti: oggi ($t = 0$), in cui compriamo o vendiamo titoli ai prezzi correnti, e domani ($t = 1$), in cui si realizza uno tra m *stati del mondo* e ogni titolo paga un importo noto che dipende dallo stato. *Decidiamo* le quantità di ciascun titolo in portafoglio (positive se comprate, negative se vendute allo scoperto). *L'obiettivo* è minimizzare il costo del portafoglio oggi. *I vincoli* impongono che domani, in *ogni* stato del mondo, il portafoglio non perda: se con queste regole si riesce a spendere meno di zero (cioè a incassare), quello è un arbitraggio.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di titoli rischiosi; i titoli sono numerati $0, 1, \ldots, n$ e il titolo 0 è privo di rischio
m	$\in \mathbb{Z}_{\geq 1}$	numero di stati del mondo a $t = 1$
r	$\in \mathbb{Q}_{\geq 0}$	tasso privo di rischio; $R = 1 + r$ è il rendimento lordo del titolo 0, uguale in ogni stato
s_i^0	$\in \mathbb{Q}_{>0}$	prezzo del titolo i oggi ($s_0^0 = 1$: il titolo 0 costa 1 e paga R)
s_{ij}^1	$\in \mathbb{Q}_{\geq 0}$	payoff del titolo i nello stato j ($s_{0j}^1 = R$ per ogni stato j)

Variabili decisionali

Introduciamo una variabile libera per ciascuno degli $n + 1$ titoli:

$$x_i = \text{quantità del titolo } i \text{ in portafoglio} \quad (> 0 \text{ acquisto}, < 0 \text{ vendita allo scoperto}),$$
$$\forall i \in \{0, 1, \ldots, n\}.$$

Usando queste variabili, il modello LP per la ricerca di un arbitraggio è il seguente:

Modello LP di ricerca dell'arbitraggio

$$\min \quad \sum_{i=0}^{n} s_i^0 \, x_i \tag{9.1a}$$

$$\text{soggetto a} \quad \sum_{i=0}^{n} s_{ij}^1 \, x_i \geq 0, \qquad \forall j \in \{1, 2, \ldots, m\}, \tag{9.1b}$$

$$x_i \gtreqless 0, \qquad \forall i \in \{0, 1, \ldots, n\}. \tag{9.1c}$$

Descrizione della funzione obiettivo e dei vincoli del modello:

- la funzione obiettivo lineare (9.1a) minimizza il costo del portafoglio oggi: un valore negativo è un incasso immediato;
- i vincoli lineari (9.1b) impongono che domani il portafoglio non perda in nessuno degli m stati del mondo (m vincoli lineari);
- i vincoli (9.1c) definiscono le variabili, libere: si può comprare e vendere allo scoperto ($n + 1$ variabili).

Leggere l'esito: tipo A, tipo B, e la normalizzazione

Il portafoglio nullo è sempre ammissibile, quindi l'ottimo è sempre ≤ 0. Tre esiti possibili:

- **ottimo** < 0 (*arbitraggio di tipo A*): si incassa oggi senza rischi domani. I vincoli sono omogenei: la strategia si può scalare a piacere e il modello è *illimitato*. Per ottenere un certificato finito si aggiunge la *normalizzazione* $\sum_{i=0}^{n} s_i^0 \, x_i \geq -1$ ("incasso oggi al più 1"): l'ottimo diventa -1 e la soluzione è una strategia concreta;
- **ottimo** $= 0$ **con qualche vincolo di stato lasco in ogni soluzione ottima** (*arbitraggio di tipo B*): oggi non costa nulla, domani non si perde mai e in almeno uno stato si guadagna;
- **ottimo** $= 0$ **e all'ottimo tutti i vincoli attivi**: nessun arbitraggio, i prezzi sono *coerenti*.

9.3 La dualità è la prezzatura: le probabilità neutrali al rischio

Applichiamo al modello (9.1) le regole della dualità del capitolo 2 (primale di minimo: tabella da destra a sinistra). Ogni vincolo di stato (9.1b), di verso $\geq$, genera una variabile duale $p_j \geq 0$; ogni variabile libera x_i genera un vincolo duale di uguaglianza; i termini noti (tutti nulli) diventano l'obiettivo del duale:

Il duale: misure di prezzatura

$$\max \quad \sum_{j=1}^{m} 0 \cdot p_j \tag{9.2a}$$

$$\text{soggetto a} \quad \sum_{j=1}^{m} s_{ij}^1 \, p_j = s_i^0, \qquad \forall i \in \{0, 1, \ldots, n\}, \tag{9.2b}$$

$$p_j \geq 0, \qquad \forall j \in \{1, 2, \ldots, m\}. \tag{9.2c}$$

- l'obiettivo (9.2a) è identicamente nullo: del duale conta solo l'*ammissibilità* (ogni soluzione ammissibile è ottima);
- i vincoli (9.2b) dicono che il prezzo di oggi di ogni titolo è la somma dei suoi payoff pesati con i p_j: i duali *prezzano* i titoli ($n+1$ vincoli lineari);
- i vincoli (9.2c) definiscono le variabili (m variabili non negative).

Il vincolo duale del titolo 0 dà $\sum_{j=1}^{m} R p_j = 1$, cioè $\sum_{j=1}^{m} p_j = 1/R$: i numeri $q_j = R p_j$ sono non negativi e sommano a 1 — sono *probabilità*, dette **neutrali al rischio**. Con esse i vincoli (9.2b) si leggono:

$$s_i^0 = \frac{1}{R} \sum_{j=1}^{m} q_j \, s_{ij}^1, \qquad \forall i \in \{0, 1, \ldots, n\} :$$

il prezzo di oggi è il valore atteso scontato dei payoff di domani, sotto le probabilità q.

Teorema fondamentale della prezzatura

Non esiste alcun arbitraggio (di tipo A né di tipo B) *se e solo se* esiste una misura di probabilità neutrale al rischio con $q_j > 0$ per ogni $j \in \{1, 2, \dots, m\}$.

Perché è vero: dualità forte e scarti complementari

È il capitolo 2 al lavoro. Se non c'è arbitraggio di tipo A, l'ottimo del primale è 0: per la *dualità forte* il duale è ammissibile, cioè una misura di prezzatura $p \geq 0$ esiste. Se non c'è nemmeno arbitraggio di tipo B, tutti i vincoli di stato sono attivi all'ottimo, e per gli *scarti complementari* esiste una soluzione duale con $p_j > 0$ per ogni stato: moltiplicando per R si ottengono le $q_j > 0$. Viceversa, se esiste $q > 0$, ogni strategia con payoff non negativi ha costo $\sum_i s_i^0 x_i = \frac{1}{R} \sum_j q_j \left(\sum_i s_{ij}^1 x_i\right) \geq 0$: nessun arbitraggio.

9.4 Esempio svolto a mano

Tre stati, un arbitraggio evidente

Mercato con $m = 3$ stati, $r = 4\%$ ($R = 1{,}04$) e due titoli rischiosi con payoff $s^1_{1\cdot} = (10, 15, 13)$ e $s^1_{2\cdot} = (30, 15, 25)$; prezzi di oggi $s^0_1 = 6$ e $s^0_2 = 20$.

- **Passo 1 — una strategia candidata.** Compriamo un'unità di ciascun titolo rischioso ($x_1 = x_2 = 1$, spesa 26) vendendo allo scoperto il titolo privo di rischio per 27 ($x_0 = -27$). Costo oggi:

  $$1 \cdot (-27) + 6 \cdot 1 + 20 \cdot 1 = -1 :$$

  incassiamo 1.

- **Passo 2 — nessun rischio domani.** Nei tre stati il portafoglio paga

  $$-27 \cdot 1{,}04 + 10 + 30 = 11{,}92, \quad -27 \cdot 1{,}04 + 15 + 15 = 1{,}92, \quad -27 \cdot 1{,}04 + 13 + 25 = 9{,}92,$$

  tutti ≥ 0: è un arbitraggio di tipo A.

- **Passo 3 — perché il modello è illimitato.** Raddoppiando la strategia si incassano 2 oggi con payoff ancora ≥ 0: senza la normalizzazione il costo può scendere quanto si vuole.

9.5 Caso di studio

Stesso mercato dell'esempio ($R = 1{,}04$, payoff in `dati/arbitraggio_payoff.csv`), analizzato con lo script in due configurazioni di prezzi: quella con arbitraggio ($s^0_1 = 6$, $s^0_2 = 20$) e una coerente ($s^0_1 = 13$, $s^0_2 = 18{,}6923$).

```
x = m.addVars(n1, lb=-GRB.INFINITY, name="x")    # posizioni libere
v_stato = m.addConstrs(
    (gp.quicksum(s1[i, j] * x[i] for i in range(n1)) >= 0
     for j in range(n_stati)), name="stato")
if normalizza:    # senza: con arbitraggio il modello e' illimitato
    m.addConstr(gp.quicksum(s0[i] * x[i] for i in range(n1)) >= -1)
```

```
m.setObjective(gp.quicksum(s0[i] * x[i] for i in range(n1)),
               GRB.MINIMIZE)
```

9.6 Risultati

```
Prezzi (6, 20):      senza normalizzazione: UNBOUNDED (tipo A)
  normalizzato: ottimo -1  ->  incasso 1 oggi, payoff (0; 0; 0,20) >= 0
Prezzi (13, 18.69):  ottimo 0, nessun arbitraggio
  duali dei vincoli di stato p* = (0,2846; 0,6769; 0)
  q = R p*             = (0,296;  0,704;  0)   somma 1
  verifica prezzatura: s0 = (1; 13; 18,6923) riprodotti esattamente
```

I duali prezzano

Con i prezzi coerenti l'LP restituisce ottimo 0 e, nei duali dei vincoli di stato, la misura di prezzatura: $q = (0{,}296;\ 0{,}704;\ 0)$. Ogni titolo — e ogni titolo *nuovo* — si prezza come valore atteso scontato dei payoff sotto q. La $q_3 = 0$ segnala una degenerazione informativa: lo stato 3 non serve a prezzare questi titoli (il suo payoff è replicabile con gli altri due stati), ed è il motivo per cui qui il "se e solo se" del teorema richiede $q > 0$: con $q_3 = 0$ resta un arbitraggio di tipo B latente sullo stato 3.

9.7 Analisi di sensitività: l'intervallo dei prezzi coerenti

Se sono quotati solo il titolo 0 (prezzo 1) e il titolo 1 (prezzo 13), quali prezzi del titolo 2 *non* creano arbitraggio? Due LP sulle misure di prezzatura — minimizzare e massimizzare $\sum_j s^1_{2j}\, p_j$ sui $p \geq 0$ coerenti con i titoli quotati — danno l'intervallo

$$s^0_2 \in [18{,}69,\ 21{,}54].$$

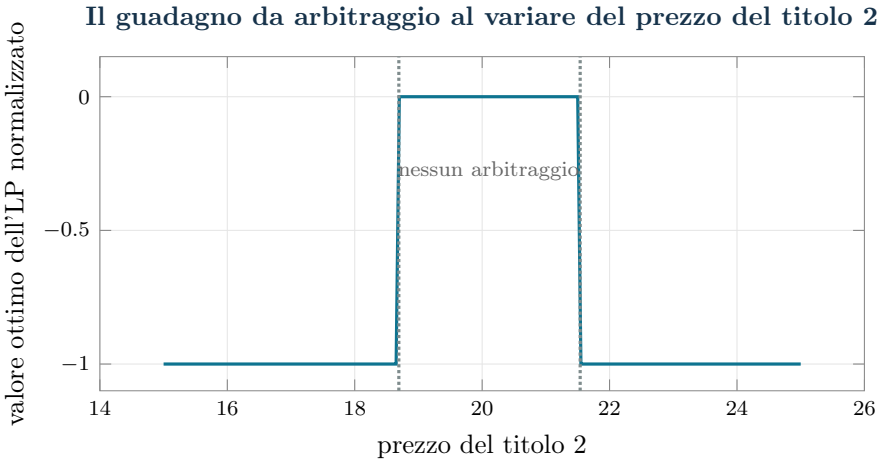


Figura 9.1: Valore ottimo dell'LP di arbitraggio normalizzato al variare del prezzo del titolo 2 (titoli 0 e 1 quotati a 1 e 13). Dentro l'intervallo [18,69, 21,54] (linee tratteggiate) il valore è 0: prezzi coerenti; fuori, il valore scende a -1: esiste un arbitraggio, in una direzione o nell'altra.

Prezzare un titolo nuovo: completo vs incompleto

La stessa coppia di LP prezza qualunque titolo nuovo. Una *call* sul titolo 1 con strike 12 paga $\max(s^1_{1j} - 12, 0) = (0,\ 3,\ 1)$ nei tre stati.

- **Mercato completo** (titoli 0, 1, 2 quotati): la misura di prezzatura è unica e il prezzo pure: 2,0308.
- **Mercato incompleto** (solo titoli 0 e 1): le misure coerenti sono infinite e il prezzo è un *intervallo*, [1,4615, 2,0308]: sotto il minimo o sopra il massimo, chiunque potrebbe costruire un arbitraggio combinando call e titoli quotati.

9.8 In sintesi

In sintesi

- Scoprire un arbitraggio è un LP con posizioni libere e un vincolo per stato: payoff ≥ 0.
- Ottimo < 0 è un arbitraggio di tipo A, la normalizzazione (costo ≥ -1) evita l'illimitatezza; ottimo 0 con un vincolo di stato non attivo è il tipo B.
- I duali, moltiplicati per R, sono probabilità neutrali al rischio: niente arbitraggio se e solo se esiste q con $q_j > 0$ per ogni stato.
- Due LP danno l'intervallo dei prezzi coerenti: un prezzo unico in un mercato completo, un intervallo in uno incompleto.

9.9 Esercizi

Esercizio 9.1 — scala dell'arbitraggio

Con i prezzi $(6, 20)$, trovare una strategia che incassa 5 oggi senza rischi domani. (Suggerimento: le soluzioni dell'LP omogeneo si possono scalare.)

Esercizio 9.2 — estremo dell'intervallo

Porre il prezzo del titolo 2 all'estremo superiore 21,5385 e risolvere l'LP di arbitraggio: verificare che l'ottimo è 0 e calcolare la misura di prezzatura associata. Che cosa succede appena sopra, a 21,60?

Esercizio 9.3 — prezzatura di una call

Nel mercato completo (prezzi 1, 13, 18,6923), prezzare una call sul titolo 2 con strike 20 (payoff $\max(s^1_{2j} - 20, 0)$) e verificare il prezzo con le probabilità neutrali al rischio del caso di studio.

Parte IV

Ottimizzazione e machine learning

Capitolo 10

Regressione robusta e quantile: stimare i parametri con un LP

Classe: LP **Script:** `python/lab10_regressione.py`
Pagina online

Prerequisiti

- Da questo modulo, Capitolo 7 «Il Newsvendor e le sue varianti»: la regola del quantile e il frattile critico.
- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: il duale di un LP e i prezzi ombra.

Obiettivi

Alla fine di questo capitolo sarai in grado di:

- formulare la regressione con scarti assoluti come LP;
- leggerne il duale, riconoscendo i punti di appoggio della stima e la proprietà del quantile;
- stimare la quantità da produrre condizionata alle condizioni del giorno;
- selezionare le variabili esplicative rilevanti con un budget sui coefficienti, leggendone il prezzo ombra.

10.1 Motivazione gestionale

Tutti i modelli visti finora partono da parametri noti: i costi del capitolo 4, i prezzi orari del capitolo 5, gli scenari del capitolo 7. Quei numeri, però, qualcuno li deve stimare — e la stima è essa stessa un problema di ottimizzazione, che si risolve con lo stesso solver del resto del corso.

Qui usiamo l'ottimizzazione per *prevedere una quantità*, e la scelta della funzione obiettivo diventa una decisione gestionale: minimizzando la somma degli scarti *assoluti* il problema resta un LP, la stima non si lascia trascinare dai giorni anomali e — pesando in modo asimmetrico gli scarti in eccesso e in difetto — si può puntare a un *quantile* della domanda invece che al suo centro. È esattamente il numero che serve al newsvendor del capitolo 7.

10.2 Il modello

Il problema a parole. *Decidiamo* i coefficienti di una funzione lineare che, a partire dalle caratteristiche osservate di un giorno, prevede la domanda. *L'obiettivo* è rendere piccoli gli scostamenti fra domanda osservata e domanda prevista, pesando in modo asimmetrico quelli in eccesso e quelli in difetto. *I vincoli* definiscono, osservazione per osservazione, lo scostamento della previsione dal dato.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di osservazioni storiche, indicizzate da $i \in \{1, 2, \dots, n\}$
p	$\in \mathbb{Z}_{\geq 0}$	numero di caratteristiche osservate, indicizzate da $j \in \{1, 2, \dots, p\}$
x_{ij}	$\in \mathbb{Q}$	valore della caratteristica j nell'osservazione i
y_i	$\in \mathbb{Q}$	valore osservato da spiegare nell'osservazione i (la domanda)
τ	$\in \mathbb{Q}$, $0 < \tau < 1$	quantile a cui si punta: $\tau = 1/2$ dà la mediana, τ grande una previsione prudente verso l'alto

Variabili decisionali

Introduciamo $p + 1$ variabili libere per la funzione da stimare,

$$\begin{cases} w_j = \text{coefficiente della caratteristica } j \\ b = \text{intercetta della funzione} \end{cases} \qquad \forall j \in \{1, 2, \dots, p\},$$

e, per ogni osservazione, due variabili non negative che misurano lo scostamento della previsione dal dato:

$$\begin{cases} u_i = \text{scostamento in eccesso dell'osservazione } i \\ v_i = \text{scostamento in difetto dell'osservazione } i \end{cases} \qquad \forall i \in \{1, 2, \dots, n\}.$$

Usando queste variabili, il modello LP di regressione quantile è il seguente:

Modello LP di regressione quantile

$$\min \quad \sum_{i=1}^{n} \big[\tau\, u_i + (1 - \tau)\, v_i \big] \tag{10.1a}$$

$$\text{soggetto a} \quad \sum_{j=1}^{p} w_j\, x_{ij} + b + u_i - v_i = y_i, \qquad \forall i \in \{1, 2, \dots, n\}, \tag{10.1b}$$

$$w_j \gtreqless 0, \qquad \forall j \in \{1, 2, \dots, p\}, \tag{10.1c}$$

$$b \gtreqless 0, \tag{10.1d}$$

$$u_i,\ v_i \geq 0, \qquad \forall i \in \{1, 2, \dots, n\}. \tag{10.1e}$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (10.1a) paga τ per ogni unità di scostamento in eccesso e $1-\tau$ per ogni unità di scostamento in difetto: con $\tau = 1/2$ è metà della somma degli scarti assoluti, con τ grande restare sotto il dato costa più che restare sopra;
- i vincoli lineari (10.1b) definiscono lo scostamento di ogni osservazione come differenza fra le sue due parti, $u_i - v_i = y_i -$ previsione (n vincoli lineari);
- i vincoli (10.1c), (10.1d) e (10.1e) definiscono le variabili del modello: coefficienti e intercetta liberi, scostamenti non negativi.

Perché due variabili per un solo scostamento

Lo scostamento $y_i - \left(\sum_j w_j x_{ij} + b\right)$ ha segno qualunque, ma l'obiettivo deve pagarne il *valore assoluto*: spezzandolo nella parte positiva u_i e nella parte negativa v_i, entrambe non negative, l'obiettivo resta lineare. All'ottimo almeno una delle due è nulla: se fossero entrambe positive, ridurle della stessa quantità lascerebbe soddisfatto il vincolo (10.1b) e farebbe diminuire l'obiettivo, perché $\tau > 0$ e $1-\tau > 0$. È lo stesso trucco delle parti positive già usato per il newsvendor a scenari nel capitolo 7.

Senza caratteristiche: torna il quantile empirico

Con $p = 0$ resta la sola intercetta e il modello sceglie un unico numero b. Sette giorni di vendite osservate: 88, 96, 104, 112, 120, 132, 148 pezzi. La panetteria del capitolo 7 ha $c_u = 9$ e $c_o = 4$, quindi $\tau = \alpha^* = 9/13 = 0{,}6923$.

- **Passo 1 — che cosa costa muovere b.** Spostando b di una unità verso l'alto si risparmiano τ su ogni osservazione che sta *sopra* b e si pagano $1-\tau$ in più su ogni osservazione che sta *sotto*. Conviene salire finché le osservazioni sopra sono tante da compensare.
- **Passo 2 — l'ottimo.** Il valore ottimo è $\tilde{b} = 120$, la quinta osservazione in ordine crescente: sotto ce ne sono 4, sopra 2, e vale $4/7 = 0{,}5714 \le \tau \le 5/7 = 0{,}7143$. Il costo ottimo è

$$\tilde{z} = \tfrac{4}{13}(32+24+16+8) + \tfrac{9}{13}(12+28) = \tfrac{320+360}{13} = \tfrac{680}{13} = 52{,}3077.$$

- **Passo 3 — i duali.** I duali dei sette vincoli valgono $\tilde{\boldsymbol{\pi}} = \left(-\tfrac{4}{13}, -\tfrac{4}{13}, -\tfrac{4}{13}, -\tfrac{4}{13}, -\tfrac{2}{13}, \tfrac{9}{13}, \tfrac{9}{13}\right)$: $\tau = 9/13$ sulle osservazioni sopra la stima, $-(1-\tau) = -4/13$ su quelle sotto, un valore interno sull'osservazione interpolata. La loro somma è nulla e il valore duale $\sum_i \tilde{\pi}_i\, y_i = 680/13$ coincide con quello primale.

Il modello, senza caratteristiche, restituisce il **quantile empirico**: è la regola del quantile del capitolo 7, scritta come LP. Con le caratteristiche diventa un quantile *condizionato* alle condizioni del giorno.

10.3 Caso di studio: lo storico della panetteria

Sessanta giorni di vendite della panetteria del capitolo 7: per ogni giorno la domanda in pezzi, la temperatura massima, il prezzo praticato e l'indicazione di weekend (`dati/panetteria_storico.csv`). Cinque di quei giorni cadono durante due scioperi dei trasporti: il centro è deserto e le vendite crollano a una ventina di pezzi.

Stimiamo prima la relazione con la sola temperatura ($p = 1$), che si vede in un grafico, minimizzando la somma degli scarti assoluti (LP, $\tau = 1/2$) due volte: su tutti i 60 giorni e sui soli 55 giorni normali.

```
LP sui 60 giorni               : domanda = 209,77 - 3,63 * temperatura
LP senza i 5 giorni di sciopero: domanda = 210,49 - 3,58 * temperatura
Scarto assoluto medio sui 55 giorni normali: 13,08 pezzi
Distanza fra le due intercette: 0,72 pezzi
```

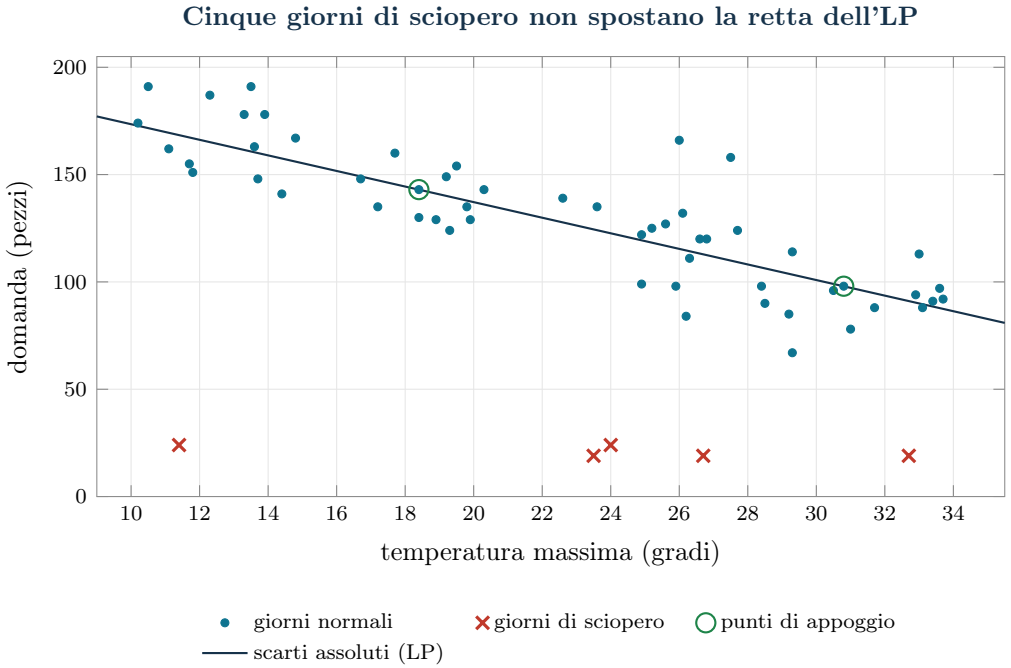


Figura 10.1: I 60 giorni della panetteria: la retta dell'LP (blu) ignora i cinque giorni di sciopero e resta a 0,72 pezzi, sull'intercetta, da quella stimata sui soli giorni normali. I due punti cerchiati in verde sono i punti di appoggio, gli unici che la retta dell'LP interpola.

> **Insight manageriale**
>
> L'LP paga ogni scarto *una volta sola*, in proporzione alla sua grandezza: cinque giorni con scarti di un centinaio di pezzi pesano come cento giorni con scarti di cinque, e la retta resta dove sono i dati normali. Il modello arriva da solo alla risposta che si avrebbe dopo aver ripulito a mano lo storico: è la differenza tra un modello che va sorvegliato e uno che si può lasciare girare ogni settimana.

Minimizzare invece la somma degli scarti *al quadrato* porta alla regressione ai minimi quadrati, un modello quadratico: il confronto con l'LP è nel modulo *Modellazione PNL*, pagina «Regressione ai minimi quadrati».

10.4 La lettura duale: i punti di appoggio

Associando un duale π_i a ciascun vincolo di scostamento (10.1b), la condizione di ottimalità sulle variabili u_i e v_i (non negative) limita ogni duale a una finestra che dipende solo da τ:

$$-(1-\tau) \;\le\; \tilde{\pi}_i \;\le\; \tau, \qquad \forall i \in \{1, 2, \ldots, n\}, \tag{10.2}$$

mentre l'ottimalità sulle variabili libere w_j e b dà le due condizioni

$$\sum_{i=1}^{n} \tilde{\pi}_i \, x_{ij} = 0 \quad \forall j \in \{1, 2, \ldots, p\}, \qquad \sum_{i=1}^{n} \tilde{\pi}_i = 0. \tag{10.3}$$

Gli scarti complementari dicono dove si trova ciascun duale:

Posizione dell'osservazione	Variabile positiva	Duale
sopra la stima (y_i maggiore)	$\tilde{u}_i > 0$	$\tilde{\pi}_i = \tau$
sotto la stima (y_i minore)	$\tilde{v}_i > 0$	$\tilde{\pi}_i = -(1-\tau)$
sulla stima (interpolata)	$\tilde{u}_i = \tilde{v}_i = 0$	$-(1-\tau) < \tilde{\pi}_i < \tau$

Le osservazioni interpolate sono i **punti di appoggio**: sono le uniche il cui duale non è incollato a un estremo, e sono loro a determinare la stima. In generale sono $p+1$, tante quante le variabili libere del modello: spostare una qualunque delle altre osservazioni, purché resti dal suo lato della stima, non cambia nulla.

```
Osservazioni sopra la retta: 29, sotto: 29, interpolate: 2  <- p + 1 = 2
pi = +0,5 (punto sopra la retta): 29 punti
pi = -0,5 (punto sotto la retta): 29 punti
somma dei duali                          = 0,000000
somma dei duali per la temperatura       = 0,000000
valore duale 616,9960 = valore primale 616,9960   (dualita' forte)
```

Il duale dell'intercetta è la proprietà del quantile

La condizione $\sum_i \tilde{\pi}_i = 0$ è il duale dell'intercetta, ed è la proprietà del quantile scritta in forma algebrica: sostituendo i valori della tabella, $\tau\, n_+ - (1-\tau)\, n_- + \sum_{i \text{ interpolate}} \tilde{\pi}_i = 0$, dove n_+ e n_- contano le osservazioni sopra e sotto la stima. A meno dei $p+1$ punti di appoggio la stima lascia quindi sotto di sé una frazione τ delle osservazioni: con $\tau = 1/2$, metà sopra e metà sotto (29 e 29 nell'esempio). La condizione $\sum_i \tilde{\pi}_i \, x_{ij} = 0$ dice la stessa cosa caratteristica per caratteristica: il segno degli scostamenti non deve avere alcun legame residuo con i dati.

Insight manageriale

Pochissime osservazioni, i punti di appoggio, sostengono la soluzione: le altre potrebbero sparire senza cambiarla, purché restino dal loro lato della stima. È la complementarietà del capitolo 2 al lavoro.

Lo scarto stesso è uno strumento diagnostico: stimando il modello con tutte e tre le caratteristiche sui 60 giorni, i giorni normali hanno scarto al più 23,2 pezzi e i cinque giorni di sciopero almeno 67,5. Una soglia a 40 pezzi li isola esattamente, senza sapere nulla degli scioperi.

```
Modello a tre caratteristiche sui 60 giorni:
  domanda = 276,46 - 3,97*temperatura - 23,12*prezzo + 28,51*weekend
Scarto assoluto: 23,2 pezzi al massimo nei giorni normali, 67,5 al minimo negli
    anomali
Giorni con scarto oltre 40 pezzi: [12, 23, 30, 42, 47] -> sono i giorni di
   sciopero
```

10.5 Regressione quantile e scorta di sicurezza

Sui 55 giorni normali stimiamo il modello a tre caratteristiche per diversi valori di τ. L'ultima colonna verifica la proprietà del quantile: la frazione di osservazioni che cade sotto la stima.

τ	intercetta	temperatura	prezzo	weekend	quota sotto la stima
0,10	269,29	−4,04	−24,71	+37,05	0,073
0,50	272,01	−3,85	−22,24	+28,74	0,455
$9/13 = 0{,}6923$	281,19	−3,96	−23,21	+31,47	0,673
0,90	287,49	−3,57	−25,69	+27,98	0,873

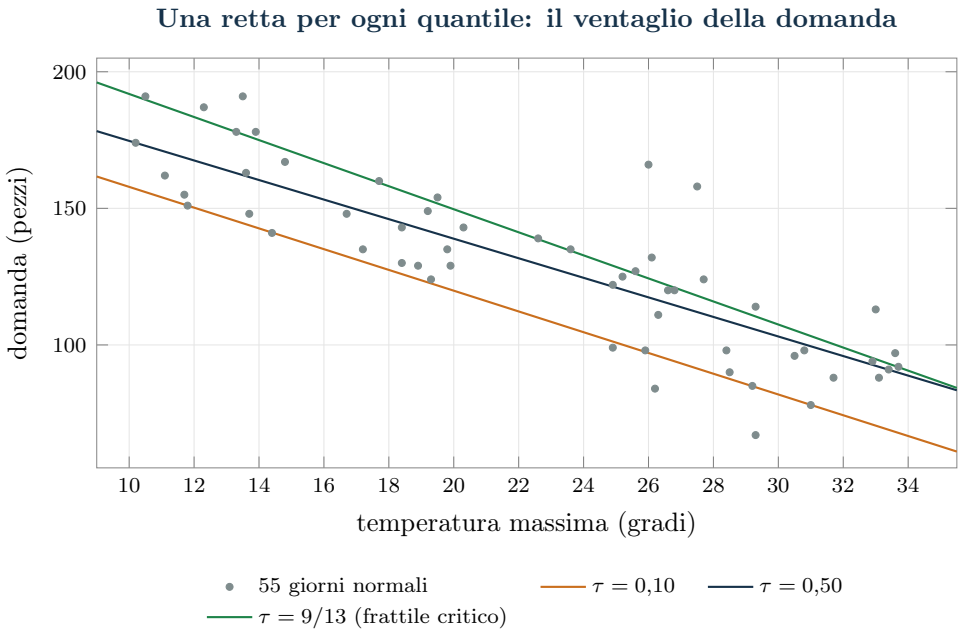


Figura 10.2: Le rette stimate con la sola temperatura per tre valori di τ: ciascuna lascia sotto di sé la frazione corrispondente di osservazioni. La retta verde, con τ pari al frattile critico della panetteria, è la quantità da produrre giorno per giorno.

Per un giorno caldo di weekend a prezzo pieno (28 gradi, prezzo 3,00) il modello dà due numeri diversi a seconda di τ:

```
domanda mediana prevista            126,3 pezzi
quantile 0,6923 previsto            132,1 pezzi   <- quantita' da produrre
scorta di sicurezza                   5,8 pezzi
quantile 0,6923 senza caratteristiche 148,0 pezzi
```

> **Insight manageriale**
>
> La regola del quantile del capitolo 7 resta valida, ma il numero a cui si applica cambia: il quantile calcolato su tutto lo storico dice di produrre 148 pezzi, quello *condizionato* alle condizioni del giorno ne chiede 132. Sedici pezzi al giorno di invenduto in meno, senza toccare i costi né la regola decisionale: solo usando l'informazione che si ha già. La differenza fra la stima mediana e quella al frattile critico, 5,8 pezzi, è la scorta di sicurezza — e il modello la calcola direttamente, senza passare da ipotesi sulla distribuzione.

10.6 Selezione delle caratteristiche con un budget

Aggiungendo caratteristiche il modello migliora sempre sui dati storici, anche quando sono irrilevanti: con abbastanza libertà finisce per inseguire il rumore. Un modo per governare il fenomeno è mettere un *budget* sui coefficienti. Servono p nuove variabili non negative,

$$z_j = \text{budget di coefficiente assegnato alla caratteristica } j, \qquad \forall j \in \{1, 2, \dots, p\},$$

e il dato $t \in \mathbb{Q}_{\geq 0}$, budget complessivo. Al modello (10.1) si aggiungono i vincoli

$$-z_j \leq w_j \leq z_j, \qquad \forall j \in \{1, 2, \dots, p\}, \tag{10.4a}$$

$$\sum_{j=1}^{p} z_j \leq t, \tag{10.4b}$$

$$z_j \geq 0, \qquad \forall j \in \{1, 2, \dots, p\}. \tag{10.4c}$$

I vincoli (10.4a) costringono z_j a essere almeno il valore assoluto di w_j (all'ottimo vale esattamente quello, perché z_j consuma budget); il vincolo (10.4b) distribuisce il budget fra le caratteristiche. Il modello resta un LP.

> Il budget somma coefficienti di caratteristiche misurate in unità diverse: gradi, euro, indicatori 0/1. Senza standardizzare, il vincolo (10.4b) confronterebbe grandezze incommensurabili e la selezione dipenderebbe dalle unità di misura. Qui ogni caratteristica è portata a media nulla e deviazione standard unitaria, quindi ogni coefficiente si legge come *pezzi per deviazione standard* della caratteristica.

Alle tre caratteristiche vere se ne aggiungono quattro senza alcun legame con la domanda (follower sui social, pioggia in una regione vicina, indice di borsa, giorno del mese). Al crescere del budget si osserva quali entrano.

```
budget  temperatura  prezzo  weekend  follower  pioggia  borsa  giorno_del_mese
    10       -10,00    0,00     0,00      0,00     0,00   0,00         0,00
    20       -18,16   -1,51     0,00      0,00    -0,33   0,00         0,00
    30       -22,43   -4,45     2,35      0,00    -0,77   0,00         0,00
    45       -26,98   -8,22     9,09      0,00    -0,63   0,00        -0,09

budget  5: scarto medio 22,37 pezzi, prezzo ombra -19,952 pezzi per unita'
budget 10: scarto medio 18,83 pezzi, prezzo ombra -18,665 pezzi per unita'
budget 20: scarto medio 13,84 pezzi, prezzo ombra -10,658 pezzi per unita'
budget 30: scarto medio 10,63 pezzi, prezzo ombra  -7,271 pezzi per unita'
budget 45: scarto medio  7,19 pezzi, prezzo ombra  -4,197 pezzi per unita'
```

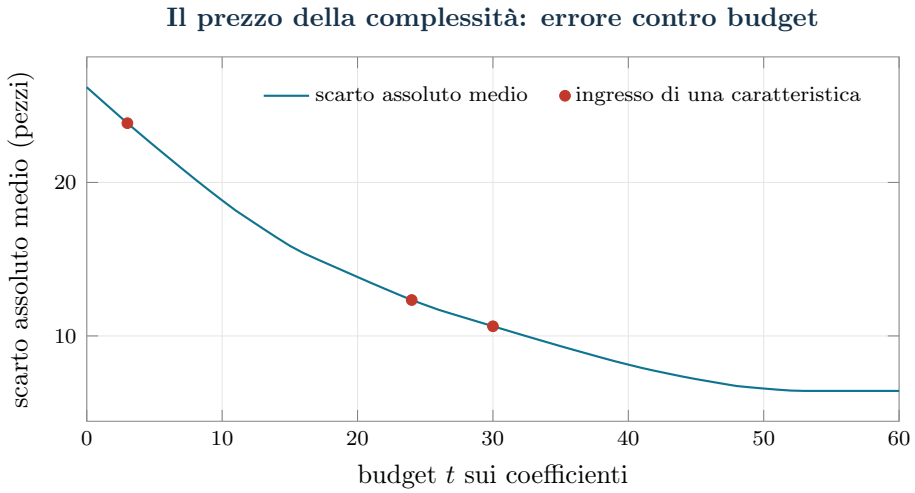


Figura 10.3: Lo scarto assoluto medio in funzione del budget sui coefficienti: una curva convessa e decrescente, come il valore ottimo di un LP al variare del termine noto. I punti rossi segnano il budget a cui entrano, nell'ordine, temperatura ($t = 3$), prezzo ($t = 24$) e weekend ($t = 30$).

Insight manageriale

Il prezzo ombra del vincolo (10.4b) è il **prezzo della complessità**: quanti pezzi di errore si risparmiano per ogni unità di budget in più. Vale $-19{,}95$ pezzi per unità quando il budget è scarso e $-4{,}20$ quando è abbondante: le prime unità comprano la temperatura, che spiega quasi tutto; le ultime comprano rumore. Le tre caratteristiche vere entrano nell'ordine della loro importanza — temperatura, prezzo, weekend — mentre nessuna delle quattro irrilevanti supera i 2 pezzi per deviazione standard finché il budget non diventa palesemente eccessivo. Il budget da scegliere non è quello che minimizza l'errore storico, ma quello oltre il quale il prezzo ombra diventa trascurabile.

10.7 In sintesi

In sintesi

- Con gli scarti assoluti la regressione è un LP: scarto $= u_i - v_i$ con $u_i, v_i \geq 0$, e la retta resta dove stanno i dati normali.
- Pagare τ sopra e $1 - \tau$ sotto dà il quantile di livello τ; con τ pari al frattile critico dà la quantità da produrre.
- Ogni duale sta in $[-(1 - \tau), \tau]$; le osservazioni interpolate sono i punti di appoggio che determinano la stima.
- Un budget su $\sum_j |w_j|$ seleziona le variabili esplicative e resta un LP; il suo prezzo ombra è il prezzo della complessità.

10.8 Esercizi

Esercizio 10.1 — quantile a mano

Con i sette giorni dell'esempio (88, 96, 104, 112, 120, 132, 148) e $\tau = 1/2$, calcolare a mano la mediana, il valore ottimo dell'obiettivo e i sette duali; verificare con lo script che la somma dei duali sia nulla.

Esercizio 10.2 — punti di appoggio

Stimare il modello con la sola temperatura e con $\tau = 1/2$, poi spostare di $+30$ pezzi un'osservazione che sta sopra la retta e rieseguire: la stima cambia? E se si sposta di -30 pezzi uno dei due punti di appoggio? Spiegare con gli scarti complementari.

Esercizio 10.3 — costo dell'errore asimmetrico

La panetteria valuta di svendere l'invenduto a fine giornata, portando c_o da 4 a 2 €. Ricalcolare α^*, rieseguire la regressione quantile e dire di quanti pezzi cambia la produzione del giorno tipo.

Esercizio 10.4 — budget e convessità

Tracciare il valore ottimo dell'LP con budget per $t \in \{0, 1, \dots, 60\}$ e verificare che sia convesso e lineare a tratti; individuare i punti di rottura e mostrare che coincidono con i budget a cui il prezzo ombra cambia valore.

Esercizio 10.5 — validazione

Dividere i 55 giorni normali in stima (40) e verifica (15). Per $t \in \{5, 10, \dots, 55\}$ tracciare lo scarto assoluto medio sui due insiemi: quello di stima è monotono in t? E quello di verifica? Quale budget si sceglie?

Parte V

Organizzazione del laboratorio

Capitolo 11

Organizzazione del laboratorio

11.1 Percorso essenziale in quattro laboratori

	Contenuto	Obiettivi di apprendimento
Lab 1	Produzione e scorte (cap. 4)	formulare un LP multiperiodale; leggere duali, slack e range; verificare un prezzo ombra per perturbazione
Lab 2	Data Envelopment Analysis (cap. 6)	costruire un LP per ogni unità; leggere la forma di inviluppo (peer e target); distinguere efficienza debole e forte
Lab 3	Newsvendor *oppure* CVaR (capp. 7–8)	formulare un LP a scenari; leggere il quantile ottimo; misurare il costo della protezione
Lab 4	Progetto a scelta	ricarica EV, arbitraggio o regressione robusta e quantile; presentazione manageriale dei risultati

I capitoli 7–10 possono sostituire i laboratori 3–4 in un'edizione orientata all'incertezza e alla stima dei parametri, oppure diventare approfondimenti per progetti individuali.

11.2 Errori ricorrenti da evitare

1. Leggere `.X` o `.Pi` senza controllare `m.Status`.
2. Dimenticare `lb=-GRB.INFINITY` sulle variabili libere (b della regressione, η del CVaR).
3. Interpretare un prezzo ombra fuori dal suo intervallo di validità (`SARHSLow/Up`).
4. Confondere il segno dei duali nei problemi di minimo (vincolo $\leq$: `Pi` ≤ 0).
5. Aggiornare il RHS di un vincolo che contiene costanti a sinistra (capitolo 5).
6. Ottimizzare un solo obiettivo quando il problema ne ha due (minimax puro senza costo).
7. Scegliere gli iperparametri (τ, λ, il budget t) guardando il test set o gli scenari usati per decidere.
8. Riportare sei cifre decimali da stime che ballano alla seconda.

11.3 Materiali del corso

dispensa	questa dispensa, in PDF nella pagina «Materiale» del sito
`python/`	uno script per capitolo; `esegui_tutti.py` rigenera dati e figure
`dati/`	tutti i dati dei casi di studio in CSV
`notebooks/`	un notebook Colab per capitolo, da aprire senza installare niente
`GUIDA_GUROBI.md`	guida autonoma al solver
slide	un gruppo di slide per capitolo, in PDF nella pagina «Materiale»

Riproducibilità: i due comandi che rifanno dati, tabelle e figure.

```
python3 -m pip install gurobipy matplotlib pandas scipy
python3 python/esegui_tutti.py
```

Di questi, `scipy` serve solo per le funzioni statistiche, non per l'ottimizzazione.