NUMERO: 842 DATA: 07/03/2014

A P P U N T I

STUDENTE: Costantino

MATERIA: Logistica di Distribuzione + casi di studio + temi d'esame

Prof. Zotteri

Logistica di Distribuzione

Cap. 1 : Supply Chain Management

La gestione della catena di fornitura, o supply chain management, è uno dei temi più rilevanti e vasti in ambito gestionale.

Sotto questa etichetta generica si collocano molti problemi anche piuttosto eterogenei, ma che interagiscono tra di loro in vari modi; la loro soluzione integrata è un prerequisito per il buon funzionamento dell'azienda.

All'interno del supply chain management si inseriscono diversi temi che tipicamente si associano alla logistica

La caratteristica più problematica della logistica è la varietà di situazioni che ci si trova ad affrontare. Purtroppo, o per fortuna, una strategia di gestione ottimale in un caso può rivelarsi disastrosa in un altro.

Cosa si intende per Logistica

La logistica ha una storia piuttosto lunga, che precede di molto la nascita dell'approccio "scientifico" al tema.

Un elemento chiave dell'evoluzione della società umana è stato il disaccoppiamento tra produzione e consumo, conseguenza della specializzazione delle persone

la logistica lega la produzione con il consumo

i primi problemi di logistica sono legati all'approvvigionamento degli eserciti

Napoleone può essere ricordato come un innovatore nel campo, in quanto un modo diverso di gestire la catena dei rifornimenti conferiva alle sue truppe una mobilità di gran lunga superiore a quella dei suoi avversari

Napoleone: "Esercito marcia sul suo stomaco"; è più importante dar da mangiare all'esercito (x evitare diserzioni) che altro

Logistica dell'approvvigionamento fondamentale, quindi, per tenere unito l'esercito

↳ Napoleone riusciva sempre a presentarsi con tutto il suo esercito contro i nemici, prima che questi siano riusciti a riunirsi tra

Strutture delle reti logistiche

Da un punto di vista fisico, una supply chain consiste di più stadi presso i quali gli items vengono prodotti, trasformati, assemblati, imballati e distribuiti ai consumatori.

↓

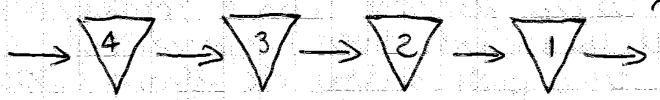
Una rete logistica è costituita da più nodi, sui quali fluiscono più tipi di prodotti. Spesso si usa il termine item o Stock Keeping Unit (SKU) per indicare in modo generico una materia prima, un componente o un prodotto finito

↓

A differenza di PCP, Logistica vede questi nodi (magazzini, fabbriche...) come delle black box contenenti delle scorte, e non come tali scorte vengono prodotte.

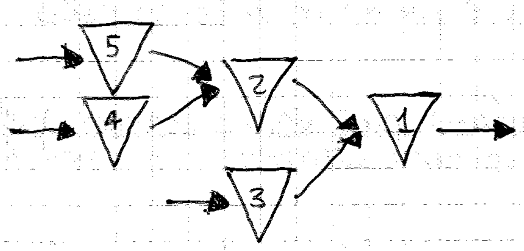
↳ l'unica cosa rilevante dal punto di vista logistico è che tali nodi hanno:
- scorte,
- lead time
- costi

• Struttura Lineare : è la struttura più semplice, costituita da una serie di mg (nodi), in cui ogni mg fornisce al più 1 mg ed è rifornito da al più 1 mg



↳ strutture molto rare

• Struttura Convergente : tipica struttura dei mg di produzione (Assemblaggio) dove ciascun mg ha al più 1 mg a valle ma può avere più mg di rifornimento (a monte)



↳ In questa situazione si pongono problemi aggiuntivi come la sincronizzazione di consegne e disponibilità per i diversi componenti da assemblare

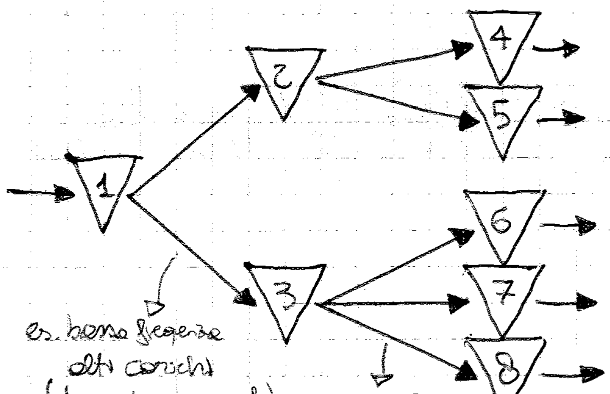
• Struttura Divergente (o Arborescente) : tipica struttura distributiva, nella quale ciascun mg ha al più 1 fornitore ma può avere più clienti

↓

In qst contesto, gli item che fluiscono sulla rete possono essere gli stessi ai diversi livelli

↓

Qui si possono porre problemi che non si incontrano nelle due strutture precedenti, come l'allocazione di scorte ai nodi successori in situazioni di mancanza di materiali



es. bassa frequenza alti carichi (domanda aggregata)

es. alta frequenza bassi carichi

Fattori di Competizione

Quando si gestisce una supply chain, lo scopo naturale è quello di fornire al cliente un buon servizio ed un costo appropriatamente basso.

Con "buon servizio", si intende che il cliente dovrebbe ricevere ciò che vuole, quando vuole e come vuole

Non esiste però una singola strategia dominante x raggiungere tale scopo: non vi è possibilità di essere il primo della classe su tutte le possibili dimensioni ad un costo ragionevole

È necessario fare un trade off tra tutte le dimensioni rilevanti, prioritizzando i fattori di competizione nell'ottica di definire una strategia

Nella logistica, l'obiettivo di minimizzare il costo è fondamentale, ma non è l'unico. Vi sono altri elementi che possono portare l'azienda al successo:

- Qualità
  - di conformità (conformance quality): tutti i pezzi devono essere conformi allo standard
    - qui la logistica è fondamentale xk deve garantire che il prodotto arrivi al cliente con la stessa qualità che aveva all'uscita dalla fabbrica
  - di target (target quality): ossia la qualità di progetto, la qualità che il prodotto dovrebbe avere in base al suo progetto (qualità di target Ferrari > qualità f.500)
    - in quest'ambito la logistica può far poco

La qualità non è rilevante solo in termini di prodotti, ma anche in termini di:

- Servizi: il concetto di servizio offerto al cliente può trovare declinazioni anche molto diverse:
  - Disponibilità del prodotto desiderato
    - la mancanza di servizio (es. stockouts) non si limita a produrre dei costi nel breve periodo (mancato ricavo ed eventuali penali) ma può avere degli effetti pervasivi di lungo periodo sul posizionamento dell'azienda inducendo alcuni clienti a scegliere un fornitore piuttosto che un altro

Es.
- Amazon.com : offre un catalogo vastissimo, aggregando la domanda in un unico grosso mg centrale
  ↓
  tramite internet riescono a gestire una numerosità di assortimento infinitamente più grande del più grande negozio fisico
- department store : a piano terra cosmetica, espositore di 100, tolgo gli 80 prodotti meno venduti → in teoria, dovrei mantenere il 90% dei clienti (quelli che acquistavano quegli 80 prodotti) e x giunta mi aspetterei che alcuni clienti di quel 10% che acquistavano 1 degli 80 prodotti che ho eliminato, ne acquisti 1 dei 20
  ↓
  in realtà non solo non supero il 90%, ma scenderò al di sotto del 90% xk alcuni clienti dei 20 prodotti non acquisteranno + da me xk ho meno assortimento

- Flessibilità : capacità di adattarsi, con tempi e costi ridotti, a cambiamenti e a condizioni eccezionali
  ↓
  è una capacità multidimensionale xk ci si può adattare a più cose diverse
  ↳ ad es, una supply chain flessibile può soddisfare un ordine estremamente importante in pochissimo tempo

→ vi sono differenti tipi di flessibilità in base alla variabile che genera la necessità di un cambiamento:

- di prodotto : abilità di adattare il prodotto ai bisogni del cliente
  ↳ maggiori/minori possibilità di personalizzazioni a scapito del DLT, costi, qualità di terze... (es Subway)
- all'innovazione di prodotto : capacità di gestire l'introduzione di un nuovo prodotto, ossia di gestire prodotti inizialmente non previsti, x i quali l'impianto non era stato progettato
  ↓
  x ottenere tale flessibilità, l'azienda dovrebbe acquistare un sistema di produzione flessibile e dovrebbe cercare di utilizzare gli stessi componenti anche x i nuovi prodotti.
  ↓
  Qst tipo di flessibilità è sempre più importante dato la crescente importanza dei nuovi prodotti e dell'innovazione
- di consegna : capacità di adattare le consegne ai bisogni dei clienti
- di mix : capacità di adattarsi a variazioni di mix (es. Hp customizzazione posticipata)

tuttavia abbiamo spesso a che fare con funzioni di costo NON lineari (possibilmente discontinue), ossia che NON variano in modo lineare col volume dell'attività (es. $f(x_1,x_2) = x_1 \cdot x_2$ o $f(x) = x^{0.6}$ ...)

↳ questo è il caso di:
- sconti qtà
- economie/diseconomie di scala
- interazioni tra le attività

Nella pratica i costi son sempre NON lineari, ma a volte possono essere opportunamente approssimati da funzioni lineari, almeno per piccole variazioni (intervalli) del livello dell'attività (tutt'al più funzioni lineari a tratti)

Si considera una generica funzione di costo $C(x)$, si definisce:

- Costo marginale $C'(x) = \frac{dC}{dx}$, ossia la derivata di 1° ordine della funzione di costo
  - ↳ costo incrementale (aggiuntivo) legato ad un'unità in più (es. qnt mi costa produrre un pezzo in più)
  - ↳ è costante per una funzione di costo lineare, ma NON in generale

- Costo medio $= \frac{C(x)}{x}$, ossia quanto mi costa in media fare uno degli x pezzi prodotti

OSS. Per funzioni di costo lineari: Costo marginale ≡ Costo medio

- Costi fissi o Variabili
  - ↳ nell'accezione contabile, con costo fisso si intende un costo affondato (sunk cost), ossia un costo che non possiamo modificare né tanto meno eliminare nel breve periodo. (costo che è stato pagato e che nessuna decisione futura può influenzare)
    - ↳ è un costo irrilevante nel breve periodo (non influenza le decisioni correnti, in quanto rappresenta una costante nella f.o. di un problema di ottimizzazione, non modificando il valore della soluzione ottima)

OSS. Nel lungo periodo tutti i costi sono variabili, pertanto la distinzione è una questione di scala temporale

Costo variabile: costo rilevante nel breve periodo, direttamente influenzato dai livelli di produzione

  - ↳ nel corso di Logistica invece, useremo i termini costi fissi e variabili

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 11 di 463

OSS. Tutti i costi sin qui considerati sono legati tra loro (al diminuire di un costo ne aumentano altri) → trade-off

per tale motivo è fondamentale rendere il sistema di rilevazione dei costi il più completo possibile (se non considero qualche costo finirò x minimizzare gli altri a discapito di questo, senza accorgermene)

Strategia

È evidente che non c'è modo di trovare un'unica soluzione che sia ottimale sotto tutti i punti di vista

In effetti, nel campo della logistica sono attivi operatori diversi, ognuno con il suo punto di vista:

- in ambito retail è fondamentale avere merce disponibile sugli scaffali. Tuttavia, un cliente insoddisfatto, non ha un impatto enorme, specialmente se non trova beni che hanno sostituti accettabili. Però uno stockout su un'intera categoria merceologica come il latte, oppure su uno specifico prodotto caratterizzato da una forte fedeltà di marca (brand loyalty) come la CocaCola potrebbe avere effetti molto più significativi
- Chi vende parti di ricambio critiche per grossi macchinari industriali, o addirittura x attrezzature ospedaliere, ha ovviamente vincoli e obiettivi diversi e, quindi, tenderà ad avere livelli di scorte più elevati e sistemi di trasporto più veloci e puntuali anche se con costi maggiori e saturazione minore

Al fine di definire una strategia, dobbiamo associare delle priorità ai fattori di competizione e trovare soluzioni efficaci ed economiche x raggiungere un dato obiettivo di performance, possibilmente facendo un trade off tra performance e costi

Identificare una strategia significa capire che in determinate situazioni alcune cose sono più importanti di altre. La rilevanza cambia, ad esempio, in base al settore in cui si opera; ma non solo:

- società in settori differenti probabilmente definiranno strategie diverse
- ma anche all'interno dello stesso settore, possiamo osservare strategie abbastanza diverse (es. Dell vs Hp, IKEA vs MondoConvenienza ...)

La strategia può quindi variare a seconda del:

- settore $<^{=}_{\neq}$
- mercato geografico (rivenditori auto UE vs USA)
- prodotti (esistono elementi chiave e livello di referenza) ↔ brand loyalty

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 13 di 463

che corrisponde ad un'assicurazione sul costo d'acquisto di tali scorte (sulla possibile fluttuazione del costo)

→ es Lavazza → mag al porto di Genova (no NA)

• **Scorte di Pipeline** (o in-transit): scorte dovute ai Lead Time

se il trasporto richiede poche ore, l'inventory in-transit è trascurabile. Nel caso invece di lunghe distanze non è più così

stessa cosa vale x i sistemi di produzione: più è lungo il flow time, più sarà grande il WIP

→ dipendono quindi da: - flusso (qtà trasportata/acquistata)
- lead time (tempi di trasporto)

Non dipendono invece dalle politiche di lottizzazione

in-transit stock medio dipende solo da - domanda media
- LT trasporto

Es: $LT_{trasp} = 1$ mese

2 politiche di lottizzazione:
(a) 1 ordine/mese di 100 pz
(b) 1 ordine/anno " 1200 pz

(a) scorte in-transit sempre di 100 pz ⇒ in media 100 pz
(b) 1° Genn partono 1200pz e arrivano il 31 Genn ⇒ scorte in-transit = 1200 a gennaio e 0 negli altri 11 mesi ⇒ in media $\frac{1200}{12} = 100$ pz

⇒ pipeline (in-transit medio) = domanda (media) · LT

x ridurle quindi ho 2 possibilità:
• ridurre domanda (poco furbo)
• ridurre LT (es. usare aereo x pezzi di valore)

• **Scorte Stagionali**: legate alla capacità

potremmo essere costretti a ricorrere a scorte di items per meglio far coincidere la domanda con la capacità

La variabilità della domanda può essere considerata come un'ulteriore motivazione a mantenere delle scorte.

Tali variabilità possono essere sostanzialmente prevedibili o no. Una sorgente di variabilità abbastanza prevedibile è la STAGIONALITÀ

Nel caso di prodotti con domanda stagionale, può essere necessario costruire una scorta nel periodo di scarsa domanda, x poter soddisfare le richieste nel periodo in cui la " è alta

7

Le SS sono scorte che tengo x fronteggiare eventi imprevisti (compresa l'incertezza della domanda ma non solo)

↓

In logistica esistono diversi fattori affetti da qualche forma di incertezza.

↳ Ad es. nell'EOQ bisognerebbe tenere conto dell'incertezza di:

- domanda
- LT del fornitore

Ad un livello decisionale diverso, nel lungo periodo, si ha incertezza anche su:

- prezzi
- tassi di cambio valute
- cambiamenti nella domanda media

I fattori di incertezza cambiano in funzione dell'orizzonte temporale e del livello gerarchico al quale le decisioni vengono prese

↓

il concetto stesso di incertezza ha diverse accezioni: la più comune è quella probabilistica, che può essere modellata da variabili casuali che seguono una data distribuzione

↓

Qualunque sia la natura dell'incertezza, dobbiamo trovare qualche modo x mitigarne gli effetti negativi.

↓

Le SS svolgono questo ruolo:

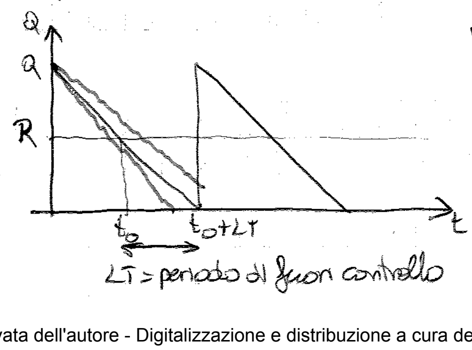
- se sia la domanda che il LT del fornitore sono costanti, è facile intuire che bisogna ordinare una qtà Q ogniqualvolta $IP \leq d \cdot LT$, ossia della domanda durante il LT
- se c'è incertezza sulla domanda e/o sul LT bisogna aumentare il livello di riordino R.

↓

x fare ciò ragionevolmente, abbiamo bisogno di:

- una descrizione dell'incertezza della domanda durante il LT
- un'adeguata definizione di una misura della qualità del servizio che vogliamo offrire al cliente (in termini di capacità di soddisfare immediatamente le loro richieste) → Service Level

Infatti, in condizioni di incertezza, dopo aver effettuato l'ordine, potremmo avere una domanda più alta o più bassa del previsto nel LT:



Pertanto, cambia quanto velocemente raggiungo il punto di riordino R

↓

ossia, cambia solo il **QUANDO** effettuare l'ordine, non quanto ordinare!!!

con questa espressione del C_{TOT}, non stiamo però considerando i costi di stockout, ma solamente quelli dovuti a cycle e safety stock

↓

Inoltre, in tale formula, sembra assente il LT, mentre intuitivamente ci si aspetta che esso abbia qualche ruolo. In effetti il LT è mascherato nella st. dev σ, perché quanto più grande è il LT di fornitura, tanto più grande è l'incertezza della domanda durante il LT

↳ le incertezze di domanda e LT si compongono nel determinare σ

↓

Idealmente, in assenza di incertezza, il LT sarebbe effettivamente irrilevante, perché comporterebbe solo un anticipo nel lancio dell'ordine

↓

Infine, tale formula, sembra suggerire che riducendo il fattore $z_{1-\alpha}$ sia possibile ottenere una riduzione di costo. Questo ci fa capire che probabilmente nella f.o. stiamo trascurando dei costi (in qst caso particolare quelli legati allo stockout)

↓

Poiché le SS servono a fare fronte a fluttuazioni di domanda e ritardi di consegna, x ridurle occorre eliminare le fonti di incertezza

↓

Per ridurre l'incertezza e quindi le SS avremo 3 possibilità:

- ↓ variabilità domanda nel LT (no promoz. → Every Day Low Prices → Walmart stores)
- migliorare previsioni (ciò riduce incertezza a parità di variabilità)
- ridurre LT (in modo da avvicinare orizzonte temporale della previsione → minor incertezza)

↓

difatti, se le domande durante diversi periodi di tempo sono v.c. INDIPENDENTI, allora si avrà:

$$\sigma_{d_{LT}} = \sqrt{LT} \cdot \sigma_d$$ (st. dev. della domanda per unità di tempo)

ossia, $\sigma_{d_{LT}}$ cresce con la radice quadrata del LT (al crescere del LT cresce l'incertezza)

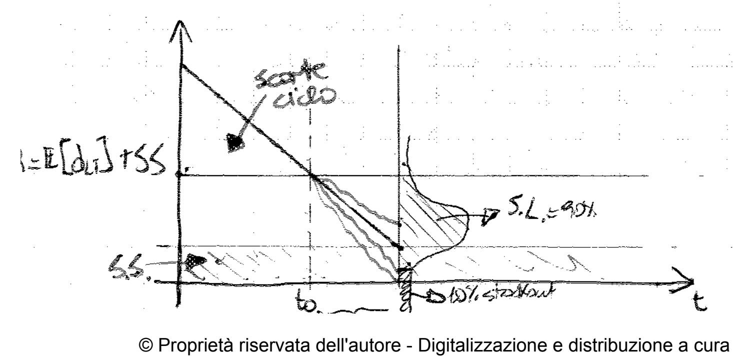
↓

N.B. tale formula è valida SOLO nel caso di NON CORRELAZIONE (indipendenza) delle domande (per uno stesso prodotto) nel tempo (ossia, domanda nel 1° periodo del LT è indipendente dalla domanda nei periodi successivi)

↓

qst xk, in assenza di COVARIANZA, si ha che

Varianza della domanda totale = Σ varianze delle singole domande



N.B. Un secondo prima di ricevere Q, in media avrò ancora SS, difatti:

$$R = \underbrace{\mathbb{E}[d_{LT}]}_{\text{ciò che mi aspetto di consumare durante LT}} + \underbrace{SS}_{\text{ciò che mi aspetto di avanzare}}$$

9

Per semplicità supponiamo di operare in un ambiente in cui esiste una finestra di vendita (tipico del newsvendor model) dopo la quale i componenti non vengono più usati

↓

Avremo pertanto 2 momenti decisionali:

1. pianificazione della produzione dei componenti (qnt unità di ogni componente produrre) sulla base dei vincoli di capacità
2. pianificazione dell'assemblaggio finale, una volta ricevuti gli ordini dei clienti, in modo da massimizzare il ricavo, soggetto a vincoli di disponibilità dei componenti

↓

L'obiettivo è massimizzare il profitto (ricavi derivanti dalla 2ª decisione meno costi associati alla 1ª), anche se non è possibile far ciò nel vero senso della parola, poiché il profitto è una v.c. che dipende dalle nostre decisioni e dalla domanda incerta. Pertanto potremo solo <u>massimizzare il valore atteso del profitto</u>

↓

Naturalmente, un approccio possibile è quello di "scommettere" su un possibile scenario di domanda, ad es. la domanda attesa

↓

Un'altra possibilità è invece quella di considerare diversi scenari possibili di domanda (risultanti dalla discretizzazione di una distribuzione di prob. continua o da un'intervista a 3 esperti)

↓

Analizziamola con un semplice esempio:

Azienda con 3 PF (j): $A_1 \rightarrow$ componenti (i) C_1, C_2, C_3; $A_2 \rightarrow$ " C_1, C_2, C_4; $A_3 \rightarrow$ " C_1, C_2, C_5 → G_{ij} = n° componenti i in J

Bill of resources (processing time T_{im})

	M_1	M_2	M_3	Costo
C_1	1	2	1	20
C_2	1	2	2	30
C_3	2	2	0	10
C_4	1	2	0	10
C_5	3	2	0	10
cap.	800	700	600	

cap. → L_m

Costo → C_i = costo componente i (costi variabili diretti di produz. + costi materiali)

Scenari di domanda (equamente probabili)

	S_1	S_2	S_3	Domanda Attesa	Prezzo Vendita
A_1	100	50	120	90	80
A_2	50	25	60	45	70
A_3	100	110	60	90	90

S_1, S_2, S_3 → d_J^s; Domanda Attesa → $\bar{d}_J$; Prezzo Vendita → prezzo P_J degli y_J (PF)

Assumiamo che Assemblaggio NON sia il bottleneck (capacità assemblaggio ∞)

La quantità assemblata è ora rappresentata dalla variabile decisionale dipendente dallo scenario y_j^s

↓

le decisioni di assemblaggio sono piani contingenti. Solo una volta che si è realizzata la domanda sceglieremo uno dei 3 piani di assemblaggio

↓

Nei modelli multiperiodali le decisioni verranno implementate in successivi periodi di tempo, ma vengono prese ora, sulla base delle info correnti. Non ci affidiamo quindi ad una specifica decisione x gli stadi successivi; la decisione che verrà implementata dipende dalla realizzazione di v.c. e sarà fissata solo quando le info rilevanti saranno disponibili

$X_1^* = 115.71$ $X_2^* = 115.71$ → piano di produzione

$X_3^* = 52.86$ $X_4^* = 2.86$ $X_5^* = 62.86$ ↳ implementato nell'immediato

$y_1^{1*} = 52.86$ $y_2^{1*} = 0$ $y_3^{1*} = 62.86$

$y_1^{2*} = 50.00$ $y_2^{2*} = 2.86$ $y_3^{2*} = 62.86$ → 3 piani di assemblaggio (contingenti)

$y_1^{3*} = 52.86$ $y_2^{3*} = 2.86$ $y_3^{3*} = 60.00$

in ogni scenario non butto mai niente di ciò che costa tanto (ho X_i sufficienti per produrre i prodotti + profittevoli)

↳ condizionali alla realizzazione di uno specifico scenario

Si nota subito che il piano è meno estremo (non si scommette più solo sulle vendite del prodotto A_3 che è caratterizzato da una domanda bassa nello scenario S_3)

↓

Notiamo inoltre che qst soluzione non differisce significativamente dalla precedente in termini di produzione dei componenti comuni

↓

NB In generale, i componenti comuni rappresentano una forma di flessibilità, xk la loro domanda deriva dalla somma delle varie domande per i diversi prodotti finiti; può accadere che la domanda per i componenti comuni abbia una variabilità inferiore a quella per i singoli prodotti finiti

↓

Aggregando la domanda, quindi, spesso si riduce l'incertezza

↳ ciò è detto Risk Pooling effect

↓

Spesso può essere conveniente impiegare componenti comuni standardizzati, anche se più costosi, proprio perchè questo permette di ridurre le scorte di sicurezza

↓

È inoltre importante notare che quando le domande dei prodotti finiti sono fortemente correlate, il risk pooling effect risulta considerevolmente ridotto

↳ in tal caso, ci si può aspettare che persino le qtà prodotte di componenti comuni differiscano tra il caso deterministico e quello stocastico

Ad es, in ambito più strettamente distributivo, alcune catene di negozi di moda, mandano all'inizio della stagione solo una quota parte dei loro prodotti nei punti vendita, riservandosi di inviare, in un secondo momento, dopo aver osservato le vendite di ogni punto vendita, la scorta residua

↓

Anche in qst caso, la 1° decisione, cioè l'acquisto dei prodotti dai fornitori, è spesso vincolata dal budget limitato a disposizione dei responsabili di ciascun segmento (buyer); mentre la 2° decisione, ossia l'allocazione delle scorte, viene presa una volta osservata parte della domanda della stagione da parte dei planner

↓

(OSS.) Nell'es precedente abbiamo scelto come f.o. il valore atteso del profitto. Qst ha senso nei casi in cui conta il profitto medio sul lungo termine. In molte situazioni può essere necessario ricorrere a diverse misure di rischio, che tengano anche conto della variabilità di profitto. Si pensi ad una situazione in cui, a fronte di perdite ingenti nell'immediato, l'azienda si trovi costretta a chiudere senza la possibilità di recuperare in futuro

Questi modelli analizzati sono single-period (nel modello non è presente il tempo)

↓

vi è un unico periodo con 2 stadi decisionali tra i quali c'è un'informazione (ossia la domanda è unica ma la decisione viene presa in 2 stadi)

Inventory Deployment (Allocazione Scorte)

I componenti comuni mitigano l'incertezza fornendo flessibilità e permettendo la posticipazione delle decisioni critiche. Questa è solo un'istanza del più generale concetto del risk pooling

↓

La scelta di dove posizionare le Safety Stock (Deployment) è tanto importante quanto quella di decidere quante scorte avere

↓

posizionare le SS a monte può ridurre il loro livello aggregato, ma al tempo stesso bisogna stare attenti ad assicurare un adeguato Service Level, che richiede di collocare le scorte a valle

↓

collocando le SS nel mg centrale avrò meno incertezza, in quanto vedo meglio dove ho gli eccessi di domanda e quindi rendo più fungibile la scorta dato che posso mandarla dove serve

↓

x fare ciò necessito di trasporti molto veloci

(Es.) azienda di muletti : ci sono vari mg che utilizzano tali muletti e bisogna rifornirli di pezzi di ricambio

↓

vi sono 3 possibili soluzioni :

Flussi informativi e Diritti Decisionali

Le info che attraversano la rete logistica e che supportano i processi di decisione e pianificazione possono essere centralizzate, cioè visibili ad un decisore unico, oppure no, nel senso che i diversi attori del sistema ne hanno una visione parziale

↓

In linea teorica, la completa disponibilità di informazioni permette ad un unico decisore di ottenere soluzioni globalmente ottime

↓

L'Information Technology (IT) rende tutto ciò una concreta possibilità, ma vi possono essere delle difficoltà insormontabili, come:

- la qualità (e affidabilità) dell'informazione: pianificatore centrale è sicuro che prenda decisioni migliori del personale del singolo supermercato che può controllare direttamente lo scaffale?

  Es. pianificatore centrale manda 100 pagnotte ad un supermercato e vede, dal sistema informativo, che ne sono stati venduti 90 alle h17, pertanto valuta che le 10 rimanenti basteranno fino a fine giornata, quando invece quelle 10 sono state rubate o sono state rovinate

  ↓

  la qualità della decisione dipende dalla qualità dell'info trasmessa

- Bullwhip effect: ad es, le info del punto di vendita (POS) possono facilmente essere raccolte e inviate al fornitore, che di conseguenza potrebbe pianificare le scorte. Allo stesso modo, potrebbero inviare al fornitore info tempestive in caso di promozioni pianificate in modo tale da evitare stockouts e perdita di clienti a causa dei picchi di domanda non previsti

  → nei fatti, un retailer che riceve una qtà di merce inferiore a quella richiesta al fornitore x mancanza di scorte, potrebbe essere tentato di ordinare più del necessario in occasione del prossimo ciclo di ordinazione come contromossa alla strategia di razionalizzazione del fornitore. Ma se quest'ultimo inviasse l'intera quantità ordinata, seguirà un periodo di bassa domanda durante il quale il retailer dovrà sbarazzarsi delle scorte in eccesso. Questo contribuirebbe ad un aumento della volatilità della domanda e del sentimento di sfiducia nei confronti del partner

  ↓

  Qst è uno dei motivi che genera l'effetto Bullwhip

  ↓

  un modo x superare tale problema sarebbe quello di centralizzare le info sulla domanda e di rendere il fornitore responsabile della gestione delle scorte

  ↓

  es. Barilla richiede ai rivenditori che comunichino non quanta merce vogliono ma quanta merce viene loro domandata, in modo da prendere lei le decisioni relative alla qtà da inviare ad ogni retailer

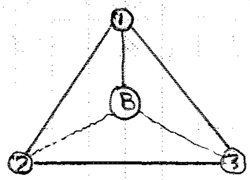
  ↓

  fece tutto ciò x ovviare al problema del nervosismo ma incontrò non poche resistenze sia interne sia esterne:

Non vi è una netta distinzione fra i 3 livelli. Essi sono interdipendenti: livelli decisionali più alti vincolano quelli più bassi, ma è valido anche in senso opposto

le scelte strategiche devono essere implementabili, pertanto nel prenderle bisogna anticipare gli effetti tattici e operativi che esse generano

Es



Network costituito da 3 clienti egualmente importanti, aventi la stessa domanda:

- se domanda è grande rispetto alle capacità dei veicoli e trasportiamo punto-punto: conviene posizionare mag in B
- se domanda è bassa e distanze non sono troppo grandi è meglio usare un modello routing x le consegne: in tal caso conviene posizionare mag lungo il perimetro o addirittura in uno dei 3 nodi (presso un cliente) ↳ (VRP)

OSS In decisioni di lungo termine, si possono preparare piani che verranno implementati in time periods successivi. Ciò può causare problemi dinamici.

Se le decisioni sono prese "qui e ora" e non vengono cambiate in seguito, si parla di problema decisionale multiperiodale ma a singolo stadio

il termine multistadio è usato x problemi nei quali le decisioni future sono adattate ad info aggiuntive raccolte progressivamente ad incertezze sempre minori

Approcci Decisionali

① Come produco/acquisto/distribuisco? Push vs Pull

- in un sistema PULL, acquisto, produzione o distribuzione degli ordini sono basati sul consumo di un bene nelle operazioni a valle. Tali politiche sono in qlk modo basate su livelli minimi di inventory e una volta che essi vengono raggiunti, lo stadio a monte inizia ad acquistare/produrre/distribuire i prodotti

- in un sistema push, acquisto, produzione o distribuzione sono basati su un piano che, a sua volta, è basato su una previsione della domanda futura

Ma in qualche modo anche le strategie pull sono basate su una sorta di previsioni. Infatti, mentre il consumo dei materiali innesca l'acquisto, la produz. o la distribuz. degli ordini, queste sono governate da parametri basati su una sorta di previsioni

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 29 di 463

Cap. 2 : Network Design and Transportation

La struttura della rete logistica si presenta in varie forme, che determinano poi vincoli sul suo modo di operare

Determinare la struttura della rete significa, in linea di principio, localizzare e dimensionare impianti di produzione, centri di distribuzione, punti vendita.

In pratica è difficile affrontare un problema in cui tutto ciò è oggetto di decisione: alcune locazioni sono tipicamente date

Ad es, nel caso di localizzazione di impianti di produzione o centri di distribuzione, si assume che siano dati i punti vendita. Viceversa, in ambito retail si pone spesso il problema di localizzare i negozi

Anche i criteri e i vincoli coinvolti saranno diversi:

Nel caso di localizzazione di negozi è fondamentale tenere conto della massima distanza, detta RAGGIO LOGISTICO, che un cliente è disposto a percorrere per acquistare un certo prodotto.

In questo caso la distanza gioca come vincolo, mentre nella localizzazione di un centro di distribuzione, la distanza tra questo e i punti vendita concorre a definire i costi di trasporto

(N.B.) Mentre nei problemi di localizzazione di impianti la domanda è assunta come nota (un dato), nel caso della localizzazione di negozi essa è influenzata dalla decisione

difatti, se il livello ultimo del network sono i negozi, la domanda sarà frutto della scelta di dove aprire il negozio e di quali dimensioni questo dovrà avere

Un negozio deve puntare ad avere più clienti possibili che abitino nel suo raggio logistico. Più è grosso il negozio, e quindi la sua offerta, più tale raggio sarà ampio e, di conseguenza, maggiore sarà la domanda

Pertanto queste scelte influiscono notevolmente sulla domanda, che quindi non possiamo considerare come data

Noi considereremo sempre la domanda come data, concentrandoci quindi sulla localizzazione di mg e centri di distribuzione, sulla ri-progettazione del network x meglio adattarsi alle mutate condizioni del mercato e sui problemi di espansione di capacità.

(1) Riduzione dell'incertezza: the Risk Pooling effect

Il costo totale per unità di tempo, ipotizzando di adottare una politica (Q,R), con lotto economico Q e punto di riordino R tale da non andare oltre una probabilità di stockout α:

$$C_{TOT} = \sqrt{2Ahd} + h z_{1-\alpha} \sigma_{d_{LT}}$$

Tale espressione ci permette di comprendere uno dei fattori che possono rendere opportuna la costruzione di un magazzino intermedio di distribuzione

Consideriamo una rete con n punti vendita, potenzialmente dotati di magazzino proprio. Per ogni punto vendita $i = 1,..,n$ abbiamo una domanda media d_i ed una deviazione standard della domanda durante il lead time σ_i. Supponiamo x semplicità che tutte le domande sui diversi punti vendita siano tra loro indipendenti e utilizziamo la formula del C_{TOT} per confrontare il costo nel caso in cui tutte le scorte vengono allocate nei punti vendita con il caso in cui si crea una struttura arborescente, con un magazzino centrale in cui sono localizzate tutte le scorte.

- Senza mg intermedio:

$$C_i = \sqrt{2Ad_ih} + h \cdot z_{LS} \sigma_i$$

$$C_{TOT} = \sum_i C_i = \sqrt{2Ah} \sum_{i=1}^{n} \sqrt{d_i} + h z_{LS} \sum_{i=1}^{n} \sigma_i$$

- Con il mg intermedio (in cui tengo tutte le SS):

$$C_{TOT} = \sqrt{2A\tilde{d}h} + h z_{LS} \tilde{\sigma}$$

incertezza della domanda del mg centrale dipende da:
- n → n° negozi
- σ_i → varianza singolo negozio
- ρ_{ij} → correlazione tra domande dei negozi

dove:

con: / senza:

ⓐ $$\sqrt{\tilde{d}} = \sqrt{\sum_{i=1}^{n} d_i} \leq \sum_{i=1}^{n} \sqrt{d_i}$$

ⓑ $$\tilde{\sigma} = \sqrt{\sum_{i=1}^{n} \sigma_i^2} \leq \sum_{i=1}^{n} \sigma_i$$

ⓐ la prima disuguaglianza suggerisce che centralizzando la gestione delle scorte si ottengono dei vantaggi dovuti ad economie di scala (mg intermedio vede una domanda maggiore e quindi i costi relativi a scorte ciclo e ordinazione saranno minori)

ⓑ la seconda disuguaglianza suggerisce che centralizzando la gestione delle scorte si ottengono dei vantaggi dovuti alla possibilità di aggregare diverse fonti di incertezza. Se le domande sono indipendenti tra loro, l'incertezza sulla domanda aggregata sarà tipicam. inferiore alla somma delle incertezze delle singole domande. Questo tipo di effetto è noto come Risk Pooling

② Varietà dei prodotti e ottimizzazione dei trasporti

Anche in assenza di incertezza, può essere opportuno introdurre nodi intermedi di distribuzione

↓

Considerando il mg del centro di distribuzione, di sicuro esso vedrà la domanda, che pure è costante sui punti vendita per ipotesi, distorta dal trasporto a lotti verso i punti vendita. Inoltre si pone la questione della sincronizzazione tra i veicoli in arrivo al centro di distribuzione e quelli in partenza, sia in termini di tempi che di quantità.

↓

Con un'opportuna sincronizzazione delle tempistiche e delle quantità, è possibile evitare di avere giacenza di mg sul centro di distribuzione che verrà, per tale motivo, chiamato transit point o nodo cross-docking

↓

Inoltre, il centro di distribuzione permette di adattare meglio il trasporto delle merce dalla produzione alla vendita. Nel caso di un trasporto punto-punto, stanti le economie di scala dei trasporti, può essere necessario trasportare una quantità eccessiva di un prodotto; peraltro, in pratica, tale quantità potrebbe essere tale da eccedere lo spazio di magazzino del punto vendita. Mescolando invece i prodotti in un punto intermedio, i veicoli possono essere gestiti in maniera più efficiente.

↓

⇒ Un centro di distribuzione può consolidare flussi di prodotti diversi in modo che questi condividano alcuni costi fissi di ordinazione e trasporto. Pertanto, ogni singolo prodotto viene spedito in quantità minori e con frequenza maggiore.

↓

In altre parole, il centro di distribuzione può creare economie di scala congiunte dal momento che i costi fissi vengono spalmati su una varietà di prodotti

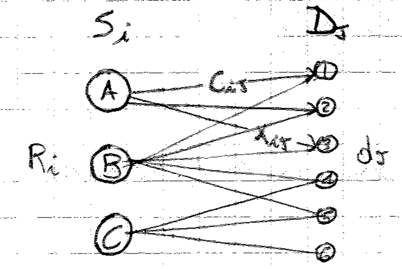
Pertanto, la scelta tra trasporto diretto o mediato dal centro intermedio dipende essenzialmente da economie di scala legate ai volumi di vendita

La nostra analisi presenta molte limitazioni, ad es:

- abbiamo considerato un solo DC, in una posizione data
- i punti vendita con volume elevato di vendite possono essere serviti direttamente dagli stabilimenti di produzione
- abbiamo ignorato i vincoli di capacità dei veicoli
- non abbiamo considerato il costo di giacenza in transit dei prodotti, che tende a penalizzare le maggiori distanze percorse nel caso di utilizzo di DC intermedi
- nel caso di DC-transit point questo non svolge una funzione in termini di risk-pooling e riduzione delle SS, ma solo una funzione di ottimizzazione dei trasporti
- material handling (x carico, scarico...) nei DC

• Il problema del trasporto:

Il classico problema del trasporto è un esempio di ottimizzazione su una rete a due livelli, su cui viaggia un solo tipo di prodotto



Vi sono 2 insiemi di nodi, le sorgenti S e le destinazioni D. Le destinazioni corrispondono ai punti vendita e sono caratterizzate da una domanda d_J nell'unità di tempo (con $J \in D$). Le sorgenti possono essere pensate come impianti di produzione, che possono fornire una qtà max R_i (con $i \in S$) nell'unità di tempo. Per ogni coppia di nodi sorgente-destinazione (i, J), abbiamo un costo unitario di trasporto C_{iJ}

Il problema consiste nel trovare il miglior modo di soddisfare la domanda, minimizzando i costi di trasporto. La decisione da prendere è quanta merce trasportare per ogni link (i,J). Indichiamo con x_{iJ} tale variabile decisionale:

$$\min \sum_{i \in S} \sum_{J \in D} c_{iJ} x_{iJ}$$

s.t.

$$\sum_{i \in S} x_{iJ} \geq d_J \quad \forall J \in D$$

$$\sum_{J \in D} x_{iJ} \leq R_i \quad \forall i \in S$$

$$x_{iJ} \geq 0$$

Il modello è estremamente semplicistico: è statico, si ignora cioè la dinamica temporale della domanda (e con essa le variabili legate alla giacenza in mg), non considera possibili costi di produz. differenziati sui diversi impianti, è monoprodotto, non considera una struttura di costi di trasporto che include costi fissi...
Facendo queste considerazioni si otterrebbe un modello integrato di produzione e distribuzione, di livello tattico-operativo

L'obiezione più rilevante alla sua struttura è legata alla linearità dei costi. Da un lato esso non rispecchia l'eventuale presenza di economie di scala, dall'altro ci troviamo di fronte ad un modello statico che in qualche modo dovrebbe tener conto di come a livello più operativo verrà trasportata la merce

• Il problema del flusso a minimo costo

Caratterizzato da una struttura di rete generica, sulla quale non sono necessariamente identificabili dei livelli. Noi ipotizziamo una rete strutturata su 3 livelli, su cui viaggiano diversi tipi di prodotti P. Non tutti i tipi di merce possono essere prodotti su tutti gli stabilimenti,

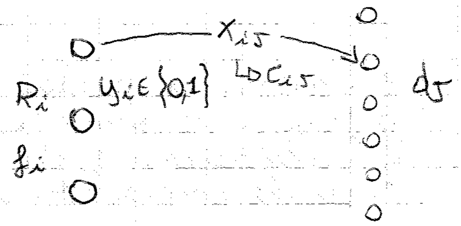
- localizzare (o rilocalizzare) gli impianti di produzione
- dimensionare (o espandere) le capacità produttive
- localizzare centri di distribuzione e determinare la loro capacità
- attribuire i diversi punti vendita ai centri di distribuzione, ovvero partizionare il mercato

Una semplice versione di questo problema è il modello di <u>plant location</u>

Esso è simile al modello di trasporto, ma in questo caso i nodi sorgente sono siti potenziali: occorre decidere quali di essi attivare (costruire), tenendo conto dei loro costi

Dobbiamo quindi introdurre una nuova variabile decisionale, binaria, legata all'apertura di un nodo sorgente:

$$y_i = \begin{cases} 1 & \text{se il nodo } i \text{ viene aperto} \\ 0 & \text{altrimenti} \end{cases}$$



In qst caso occorre anche tenere conto di un costo f_i legato all'apertura dell'impianto i (costo annuale di esercizio → costo fisso)

$$\min \sum_{i \in S} f_i \cdot y_i + \sum_{i \in S} \sum_{j \in D} c_{ij} x_{ij}$$

→ caso monoprodotto

s.t.

$$\sum_{j \in D} x_{ij} \leq R_i \cdot y_i \quad \forall i$$

$$\sum_{i \in S} x_{ij} \geq d_j \quad \forall j$$

$$x_{ij} \geq 0 \ , \ y_i \in \{0,1\}$$

Il modello di plant location si pone a livello strategico, ma ha pure una dimensione tattica legata al trasporto

Il modello presenta 2 tipi di variabili: le variabili di progetto, legate alla costruzione di impianti, e le variabili di controllo, legate all'allocazione della domanda agli stabilimenti

I 2 tipi di variabili sono modificabili su scale di tempo ben diverse, e l'opportunità di metterne in luce le differenze diventa molto più evidente se introduciamo nel modello delle incertezze sulle domande

Per fare ciò costruiamo un modello in cui sono evidenti 2 stadi decisionali, in modo da evidenziare il fatto che, prima di tutto dovremo prendere la decisione sugli impianti da aprire (scelta strategica), poi osserveremo la domanda e, in base allo scenario di domanda che si verificherà, sceglieremo le quantità x_{ij} (scelta operativa)

Supponiamo quindi di rappresentare l'incertezza introducendo un

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 39 di 463

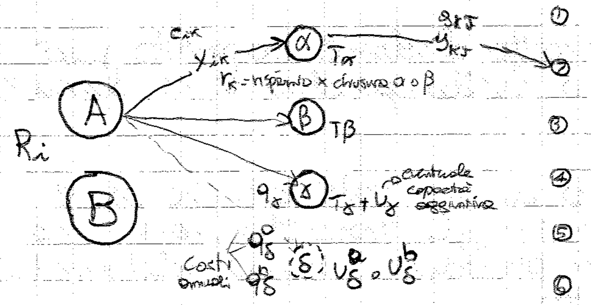
- **Riallocazione ed espansione della capacità di DC**

Al momento sono attivi 3 centri di distribuzione (DC), indicati con α, β, γ, che hanno una certa capacità T_K (con $K = \alpha, \beta, \gamma$)

Il problema è decidere come riprogettare la rete per meglio adattarsi a mutate condizioni di mercato. Vi sono le seguenti opzioni:

- espandere la capacità del centro γ
- costruire un nuovo centro δ con capacità relativamente alta / bassa

per ragioni di budget, nel caso si decidesse di aprire il nuovo centro δ, sarebbe necessario chiudere (almeno) uno tra i centri α e β (non si vogliono avere più di 3 DC attivi)



NB T_K è la capacità di gestire flussi, NON di stoccare (xk non c'è la variabile tempo)!!!

Le variabili decisionali sono:

di scenario (operative):
- i flussi di merce X_{iK}, Y_{KJ} sui due insiemi di link

di struttura (strategiche):
- $W_\gamma = \begin{cases} 1 & \text{se capacità } \gamma \text{ viene espansa} \\ 0 & \text{altrimenti} \end{cases}$
- $Z_K = \begin{cases} 1 & \text{se il centro } K = \alpha, \beta \text{ viene tenuto aperto} \\ 0 & \text{altrimenti} \end{cases}$
- $S^b_\delta = \begin{cases} 1 & \text{se si costruisce centro } \delta \text{ con capacità bassa} \\ 0 & \text{altrimenti} \end{cases}$
- $S^a_\delta = \begin{cases} 1 & \text{" " " " " alta} \\ 0 & \text{altrimenti} \end{cases}$

negativo xk non si tratta di un costo, bensì di un risparmio

$$\min \; q_\gamma \cdot W_\gamma + q^b_\delta \cdot S^b_\delta + q^a_\delta \cdot S^a_\delta - \sum_{K \in \{\alpha,\beta\}} r_K (1 - Z_K) + $$ → termini legati all'espansione di capacità

$$+ \sum_i \sum_K C_{iK} \cdot X_{iK} + \sum_K \sum_J g_{KJ} \cdot Y_{KJ}$$ → termini legati ai flussi

s.t.

$$\sum_K X_{iK} \le R_i \quad \forall i$$

$$\sum_i X_{iK} = \sum_J Y_{KJ} \quad \forall K$$

$$\sum_i X_{iK} \le T_K \cdot Z_K \quad \forall K = \alpha, \beta$$

$$\sum_i X_{iK} \le T_\gamma + U_\gamma \cdot W_\gamma$$

$$\sum_i X_{iK} \le U^b_\delta \cdot S^b_\delta + U^a_\delta \cdot S^a_\delta$$

$$S^b_\delta + S^a_\delta \le 1$$

$$Z_\alpha + Z_\beta + S^b_\delta + S^a_\delta \le 2$$

$$\sum_K Y_{KJ} \ge d_J \quad \forall J$$

$$X_{iK}, Y_{KJ} \ge 0, \quad W_\gamma, Z_\alpha, Z_\beta, S^b_\delta, S^a_\delta \in \{0,1\}$$

OSS. Se volessimo introdurre gli scenari, la f.o. diventerebbe:

$$\min \; \ldots\ldots + \sum_s \pi^s \Big(\sum_{iK} C^s_{iK} X^s_{iK} + \sum_{KJ} g^s_{KJ} \cdot Y^s_{KJ} \Big)$$

mentre i vincoli resterebbero identici, con l'unica differenza dell'aggiunta delle s in X^s_{iK}, Y^s_{KJ}, d^s_J

potenziale di molti minimi locali che richiederebbero l'uso di metodi sofisticati e onerosi per l'ottimizzazione globale

↓

Per quanto concerne gli ultimi 2 punti, è possibile affrontarli congiuntamente approssimando le funzioni NON lineari mediante funzioni lineari a tratti in cui i punti x_i demarcano le diverse zone di linearità

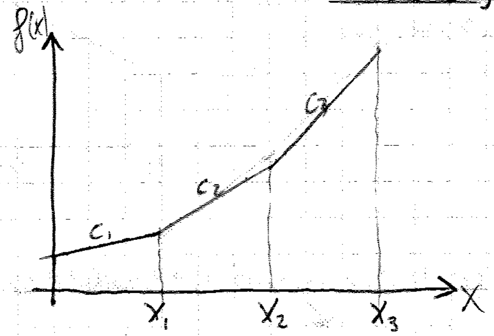
↳ sta a chi costruisce il modello decidere quanti di questi punti vanno utilizzati e come posizionarli (trade off tra complessità del modello e qualità dell'approssimazione)

In questo modo ci si riconduce, a seconda dei casi, a modelli di programmazione lineare nel continuo (se la funzione è convessa) ⓐ o a modelli misti-interi (se la funz. è concava o cmq NON convessa) ⓑ

Consideriamo ad es. una funzione che esprime il legame tra volume di produzione e costo di produzione:

$$f(x) = \begin{cases} c_1 \cdot x & , 0 \le x \le x_1 \\ c_2 (x - x_1) + c_1 x_1 & , x_1 \le x \le x_2 \\ c_3 (x - x_2) + c_2 (x_2 - x_1) + c_1 x_1 & , x_2 \le x \le x_3 \end{cases}$$

ⓐ • Se $c_1 < c_2 < c_3$, ovvero i costi marginali sono crescenti (diseconomie di scala), allora $f(x)$ è CONVESSA



↓

Il caso convesso è facile e riconducibile ad un modello di P.L. nel continuo

↓

si trasforma la funzione $f(x)$ nella somma di termini lineari, in funzione di variabili ausiliarie z_1, z_2, z_3:

$$f(x) = c_0 + c_1 \cdot z_1 + c_2 \cdot z_2 + c_3 \cdot z_3$$

$$x = z_1 + z_2 + z_3$$

$$0 \le z_1 \le x_1$$

$$0 \le z_2 \le x_2 - x_1$$

$$0 \le z_3 \le x_3 - x_2$$

In pratica, ogni variabile z_i è legata ad un intervallo (qtà di pezzi che decido di produrre in quel dato range), e poiché $c_1 < c_2$, avrò $z_2 > 0$ nella soluzione ottima solo se $z_1 \equiv x_1$. Non può accadere che all'ottimo venga usata la z_2 in luogo della z_1 (ciò non sarebbe coerente col mondo della fisica) xk questa è meno costosa.

→ Analogamente z_3 è attivata solo se sia z_1 sia z_2 sono saturate

Pertanto dovremmo imporre che:

- al più 2 lambda siano accesi (e tutti gli altri a zero)
- solo coefficienti lambda CONSECUTIVI possono assumere valori ≠ zero

↓

Quindi se è accesa la coppia λ_0, λ_1, la coppia λ_2, λ_3 sarà bloccata a zero e saremo sul primo segmento della funzione.
Analogamente, se sono liberi solo λ_1 e λ_2 saremo sul 2° segmento e così via

↓

Di conseguenza, per fare in modo che l'insieme delle mie alternative sia sulla spezzata, dovrò introdurre delle variabili decisionali binarie s_i, una per ogni segmento $(i-1, i)$ con $i = 1,2,3$, e legarle ai pesi λ_i mediante i seguenti vincoli:

$$x = \sum_{i=0}^{3} \lambda_i x_i$$

$$y = \sum_{i=0}^{3} \lambda_i y_i$$

$$0 \leq \lambda_0 \leq s_1 \longrightarrow \text{in modo che } \begin{cases} \lambda_0 = 0 & \text{solo se } s_1 = 0 \\ \lambda_0 > 0 & \text{" " } s_1 = 1 \end{cases}$$

$$0 \leq \lambda_1 \leq s_1 + s_2 \longrightarrow \text{" " "} \begin{cases} \lambda_1 = 0 & \text{" " } s_1 = s_2 = 0 \\ \lambda_1 > 0 & \text{" " } s_1 \text{ o } s_2 = 1 \end{cases}$$

$$0 \leq \lambda_2 \leq s_2 + s_3$$

$$0 \leq \lambda_3 \leq s_3$$

$$\sum_{i=1}^{3} s_i = 1 \quad , \quad s_i \in \{0,1\}$$

$$\sum_{i=0}^{3} \lambda_i = 1$$

In qst modo abbiamo compiutamente modellizzato la curva di costo indipendentemente da come sia.

↓

se ad es. l'intervallo di produzione selezionato è s_2, vorrà dire che poi, scegliendo i λ_1 e $\lambda_2 \neq 0$, il modello andrà a scegliere un volume di produzione $x = x_1 \cdot \lambda_1 + x_2 \cdot \lambda_2$ e, a fronte di questo volume di produzione, il modello ci dirà che avremo un costo di produzione pari a $y = y_1 \cdot \lambda_1 + y_2 \cdot \lambda_2$

- Se $E_1, E_2 \in \mathcal{F} \Rightarrow E_1 \cup E_2 \in \mathcal{F}$
- Se $E_1, E_2 \in \mathcal{F} \Rightarrow E_1 \cap E_2 \in \mathcal{F}$

N.B. Se $E_1 \cap E_2 = \emptyset$, i due eventi sono detti MUTUAMENTE ESCLUSIVI

MISURA DI PROBABILITÀ ($P(E)$): funzione che associa ad ogni evento $E \in \mathcal{F}$ un numero reale compreso tra 0 e 1

insieme agli altri ingredienti definisce uno:

SPAZIO DI PROBABILITÀ $(\Omega, \mathcal{F}, P)$

OSS. A parità di spazio campione si possono definire diverse famiglie di eventi

questo dipende in parte dagli scopi che ci prefiggiamo, ma anche dalle informazioni disponibili e da ciò che davvero siamo in grado di osservare, e porta a definire spazi di probabilità diversi sullo stesso spazio campione Ω

La misura di probabilità deve soddisfare le seguenti tre condizioni:

1. $0 \leq P(E) \leq 1 \quad \forall E \in \mathcal{F}$
2. $P(\Omega) = 1$
3. per ogni sequenza di eventi mutuamente esclusivi $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

Sulla base di queste proprietà basilari, si possono dimostrare tutte le proprietà che intuitivamente associamo al concetto di probabilità, ad es:

- $P(E^c) = 1 - P(E)$ poiché $E \cup E^c = \Omega$ e i due insiemi sono disgiunti, quindi: $P(E) + P(E^c) = P(\Omega) = 1$
- $(E_1 \cap E_2) = (E_1^c \cup E_2^c)^c$ per cui possiamo ragionare sulle probabilità degli eventi intersezione

PROBABILITÀ COMPOSTA ($P(E_1 \cap E_2)$ o $P(E_1, E_2)$ o $P(E_1 \cdot E_2)$)
→ probabilità che due (o più) eventi abbiano luogo congiuntamente

- nel caso di due eventi NON disgiunti $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

- Se $E_1 \subseteq E_2 \Rightarrow P(E_1) \leq P(E_2)$ ossia, se si verifica E_1 di sicuro si verifica anche E_2, ma NON viceversa

  → infatti $E_2 = E_1 \cup (E_2 \setminus E_1)$. Poiché i due insiemi sono disgiunti:

$$P(E_2) = P(E_1) + P(E_2 \setminus E_1) \Rightarrow P(E_2) - P(E_1) = P(E_2 \setminus E_1) \geq 0$$

c.v.d.

(OSS.) La conoscenza relativa ad un sottoinsieme di eventi non ci dice nulla sugli altri

→ Se 3 eventi sono indipendenti a due a due, ciò NON implica che i 3 eventi siano indipendenti

Consideriamo, ora, una partizione (finita) dello spazio campione Ω, ovvero un insieme di eventi $H_1, H_2, H_3, \dots, H_m$ mutuamente disgiunti e tali che la loro unione fornisca l'insieme Ω.
Se consideriamo un qualsiasi evento E, possiamo partizionare tale evento sulla base della partizione di Ω e considerando le intersezioni del tipo $E \cap H_i$

poiché tali intersezioni sono tra loro disgiunte abbiamo x la 3ª propr. della P(E):

$$P\left(\bigcup_{i=1}^{m}(E \cap H_i)\right) = \sum_{i=1}^{m} P(E \cap H_i)$$

Esprimendo la sommatoria in termini di probabilità condizionali, e ammettendo anche una partizione in un numero infinito numerabile di sottoinsiemi, otteniamo il seguente teorema:

TEOREMA DELLE PROBABILITÀ TOTALI: si consideri una partizione dello spazio campione Ω, cioè una famiglia di sottoinsiemi $H_1, H_2, H_3, \dots$, mutuamente esclusivi (disgiunti) e collettivamente esaustivi (tali cioè che la loro unione è pari a Ω), allora per un qualsiasi evento E, si ha:

$$P(E) = \sum_{i=1}^{\infty} P(E|H_i)\, P(H_i)$$

→ es. applicativo: $P(A) = P(A|B)\,P(B) + P(A|B^c)\,P(B^c)$

Variabili Casuali Discrete

VARIABILE CASUALE: funzione che mappa gli eventi di uno spazio di probabilità in numeri

→ si parla di VARIABILI CASUALI DISCRETE nel caso in cui i valori possibili della variabile siano:
- i numeri interi (insieme infinito ma numerabile)
- o un insieme finito (come nel caso dei dadi)

Convenzione:
- lettere maiuscole (es X) → variabili casuali (v.c.)
- " minuscole (es x) → valori assunti da una v.c. in corrispondenza della realizzazione di un evento

Come immediata conseguenza delle condizioni sulle misure di probabilità:

$$\sum_{i=1}^{\infty} p_X(x_i) = 1$$

La funzione di distribuzione cumulativa e la funzione di massa di probabilità contengono tutte le informazioni relative ad una variabile casuale discreta

Sono però anche utili delle caratteristiche numeriche che ci permettono di percepire in maniera immediata quali valori ci possiamo attendere e come essi sono dispersi (media, varianza...).

Per individuare questi parametri descrittivi, si ricorre al concetto di:

MOMENTO DI UNA VARIABILE CASUALE DISCRETA:

data una v.c. discreta X, con funzione di massa $p_X(\cdot)$, il suo momento di ordine k è definito come:

$$\mu_X^{(k)} \equiv \mathbb{E}[X^k] = \sum_{i=1}^{\infty} x_i^k \, p_X(x_i)$$

I momenti di diverso ordine permettono di descrivere le caratteristiche essenziali di una variabile casuale:

VALORE ATTESO: è il momento del primo ordine

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} p_i x_i$$

→ un importante proprietà è la sua LINEARITÀ: $\mathbb{E}[\alpha X + \beta] = \alpha \mathbb{E}[X] + \beta$

VARIANZA (σ^2): misura relativa della dispersione dei valori possibili

$$\mathrm{Var}(X) \equiv \mathbb{E}[(X - \mathbb{E}[X])^2]$$ → Varianza NON negativa per definizione

→ PROPRIETÀ:
- $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X]$ (vedi definizione)
- $\mathrm{Var}(\alpha X + \beta) = \mathbb{E}[((\alpha X + \beta) - \mathbb{E}[\alpha X + \beta])^2]$

$$= \mathbb{E}[(\alpha X + \beta - \alpha \mathbb{E}[X] - \beta)^2]$$
$$= \mathbb{E}[\alpha^2 (X - \mathbb{E}[X])^2]$$
$$= \alpha^2 \mathrm{Var}(X)$$

⇒ Varianza NON è un operatore LINEARE

Condizionando rispetto all'esito del primo esperimento è possibile ricavare agevolmente valore atteso e varianza della distribuzione geometrica

Se il primo esperimento ha successo (probabilità p) si ha $X = 1$; altrim. abbiamo già eseguito un esperimento, fallito, ed il valore atteso del numero di esperimenti ancora da eseguire x avere un successo, stante la loro indipendenza, è $\mathbb{E}[X]$, come se si ripartisse da capo:

$$\mathbb{E}[X] = \mathbb{E}[X|OK]\cdot P\{OK\} + \mathbb{E}[X|NOK]\cdot P\{NOK\} = 1\cdot p + (1+\mathbb{E}[X])(1-p)$$

↳ VALORE ATTESO: $\cancel{\mathbb{E}[X]} = \cancel{p} + 1 - \cancel{p} + \cancel{\mathbb{E}[X]} - p\mathbb{E}[X]$

$$\mathbb{E}[X] = 1/p$$

Per ricavare la varianza, troviamo in modo analogo il momento di 2° ordine:

$$\mathbb{E}[X^2] = \mathbb{E}[X^2|OK]\cdot P\{OK\} + \mathbb{E}[X^2|NOK]\cdot P\{NOK\}$$
$$= 1^2 p + \mathbb{E}[(1+X)^2](1-p) = p + (1 + 2\mathbb{E}[X] + \mathbb{E}[X^2])(1-p)$$
$$= p + (1 + 2/p)(1-p) + \mathbb{E}[X^2](1-p) = \frac{2-p}{p^2}$$

↳ VARIANZA $\quad Var(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \frac{2-p}{p^2} - \frac{1}{p^2} =$

$$= \frac{1-p}{p^2}$$

- **DISTRIBUZIONE DI POISSON**: la v.c. di Poisson è caratterizzata da un parametro λ e assume valori nell'insieme $\{0,1,2,3,\dots\}$. La funzione di massa è:

$$p_i = P\{X=i\} = e^{-\lambda}\cdot\frac{\lambda^i}{i!} \quad \text{con } i = 0,1,2,\dots$$

→ VALORE ATTESO $\quad \mathbb{E}[X] = \sum_{i=0}^{\infty} i e^{-\lambda}\frac{\lambda^i}{i!} = \lambda e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}$

$$= \lambda$$

→ VARIANZA $\quad Var(X) = \lambda \quad \Rightarrow$ Coefficiente di variazione $\quad C_X = 1/\sqrt{\lambda}$

OSS. Supponiamo di modellare il numero di clienti che arrivano nell'unità di tempo. Ogni arrivo di un cliente è un evento, ed è naturale misurare l'intensità del flusso di richieste mediante un tasso medio degli arrivi, espresso come numero medio di clienti che arrivano mediamente nell'unità di tempo. La distribuzione di Poisson è un possibile modello x tale fenomeno, adeguato quando gli arrivi

Procedendo in modo analogo al caso delle v.c. discrete, si definisce il:

MOMENTO di ordine K per una v.c. continua $\mu_X^{(k)} \equiv \mathbb{E}[X^k] = \int_{-\infty}^{+\infty} x^k f_X(x)\,dx$

Coerentemente possiamo definire:

VALORE ATTESO di una v.c. X continua $\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x)\,dx$

VALORE ATTESO DI UNA FUNZIONE g(·) di una v.c. continua $\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)\,dx$

N.B. In generale, il valore atteso della funzione NON coincide con la funzione valutata nel valore atteso, ossia:

$$\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$$

Esse coincidono nel caso particolare di una funzione lineare, stante la linearità del valore atteso, ma NON nel caso di funzioni non lineari

VARIANZA: $\mathrm{Var}(X) \equiv \int_{-\infty}^{+\infty} (x-\mathbb{E}[X])^2 f_X(x)\,dx = \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}^2[X]$

Vale la pena di notare che oltre al valore atteso esistono anche due quantità, dette moda e mediana, che caratterizzano una distribuzione di probabilità:

MODA di una distribuzione di probabilità: il punto in cui la densità $f_X(x)$ ha il suo massimo (se questo esiste)

→ viene naturale pensare che la moda sia "il valore più probabile" tra quelli assunti dalla distribuzione. In realtà, un'interpretazione del genere non ha senso nel caso continuo, in quanto tutti i valori hanno probabilità zero, mentre potrebbe sembrare più plausibile nel caso discreto, ragionando in termini di funzione di massa

OSS. Le distribuzioni teoriche più comuni sono monomodali, nel senso che la loro densità ha un solo punto di massimo locale, che è anche globale

MEDIANA: nel caso di una distribuzione continua, è quel valore m_X tale che:

$$F_X(m_X) = 0.5$$

La funzione di distribuzione cumulativa è :

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$

→ Valore Atteso $E[X] = \frac{1}{\lambda}$

→ Varianza $Var(X) = \frac{1}{\lambda^2}$

⇒ Coeff. variazione $C_X = \frac{\sigma}{\mu} = 1$

Proprietà : ASSENZA DI MEMORIA

↳ consideriamo una v.c. X distribuita esponenzialmente con parametro λ e immaginiamo che essa descriva il tempo di vita di una lampadina la cui vita media è $\frac{1}{\lambda}$

$$P(X > t) = 1 - F_X(t) = e^{-\lambda t}$$

Consideriamo la situazione che la lampadina sia ancora funzionante al tempo t e chiediamoci qual è la vita attesa residua della lampada, condizionatamente a tale informazione, ossia consideriamo la probab. che la vita della lampada superi t+s:

$$P\{X > t+s \mid X > t\} = \frac{P\{(X > t+s) \cap (X > t)\}}{P\{X > t\}} = \frac{P\{X > t+s\}}{P\{X > t\}} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P\{X > s\}$$

⇒ si nota che la vita trascorsa t NON ha nessuna influenza sulla vita residua s, il cui valore atteso è ancora $\frac{1}{\lambda}$ come se la lampadina fosse nuova

OSS. Questo modello non si presta quindi a trattare, ad es, la vita di meccanismi soggetti ad usura

una tipica applicazione della distribuzione esponenziale è x caratterizzare il tempo che passa tra l'arrivo di 2 clienti o 2 ordini successivi, detto anche tempo di interarrivo

↳ in questa interpretazione: - tasso di arrivo $= \lambda$

- Tempo medio di interarrivo $= \frac{1}{\lambda}$

Si dimostra in qst ipotesi che il numero di clienti che arriva in un intervallo di tempo t è una v.c. discreta che segue la distribuzione di Poisson con parametro λt

Dalla linearità delle operazioni di sommatoria e integrazione segue che il valore atteso di una combinazione lineare di v.c. è la C.L. dei valori attesi:

$$Z = \sum_{i=1}^{m} \lambda_i X_i \longrightarrow \mathbb{E}[Z] = \sum_{i=1}^{m} \lambda_i \mathbb{E}[X_i]$$

Questo NON vale, in generale, per la varianza:

$$Var(X+Y) \neq Var(X) + Var(Y)$$

Similmente, in generale abbiamo:

$$\mathbb{E}[g(X)h(Y)] \neq \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

N.B. Si ha l'uguaglianza solo in caso di INDIPENDENZA delle v.c.

INDIPENDENZA DI V.C.: due v.c. X e Y sono indipendenti se lo sono i due eventi $\{X \leq x\}$ e $\{Y \leq y\}$, cioè se, per qualsiasi x e y:

$$F_{X,Y}(x,y) = P\{X \leq x, Y \leq y\} = P\{X \leq x\} \cdot P\{Y \leq y\} = F_X(x) \cdot F_Y(y)$$

da questo discende che si possono fattorizzare le funzioni di massa e densità:

$$p_{X,Y}(x,y) = p_X(x)\, p_Y(y) \qquad f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$$

Questo permette di scomporre le sommatorie e gli integrali doppi, ottenendo:

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)]$$

Da questa proprietà, segue anche che nel caso di un insieme di v.c. indipendenti:

$$Var\left(\sum_{i=1}^{m} \lambda_i X_i\right) = \sum_{i=1}^{m} \lambda_i^2\, Var(X_i)$$ (valida solo se X_i sn INDIPEND.)

COVARIANZA: nel caso in cui vi sia un certo grado di dipendenza tra v.c., è naturale porsi l'obiettivo di misurarlo e se possibile sfruttarlo in qualche modo. Una semplice misura di dipendenza è la covarianza:

$$Cov(X,Y) \equiv \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

OSS. Si ha una covarianza positiva quando gli eventi $\{X > \mathbb{E}[X]\}$ e $\{Y > \mathbb{E}[Y]\}$ tendono a verificarsi insieme e la stessa cosa accade per gli eventi $\{X < \mathbb{E}[X]\}$ e $\{Y < \mathbb{E}[Y]\}$

Se invece i segni delle differenze tendono ad essere diversi si avrà una covarianza negativa, ovvero una variabile tende ad assumere valori superiori alla media quando l'altra assume valori sotto la media

Distribuzioni derivate dalla distribuzione Normale:

In generale, sommando v.c. identicamente distribuite, non si ottiene una v.c. con le medesime caratteristiche

Una notevole eccezione è la distribuzione normale:

Sommando v.c. Normali si ottiene ancora una variabile con distribuzione normale

Consideriamo un insieme di v.c. normali indipendenti X_i (con $i = 1, \dots, n$) con parametri μ_i e σ_i e costruiamo una loro C.L. con pesi λ_i:

$$X = \sum_{i=1}^{n} \lambda_i X_i$$

Dalle proprietà della somma di v.c. indipendenti sappiamo che:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \lambda_i \mu_i \qquad Var(X) = \sum_{i=1}^{n} \lambda_i^2 \sigma_i^2$$

Nel caso di v.c correlate, se raggruppiamo le covarianze in una matrice Σ (MATRICE DI COVARIANZA), i cui elementi sono $\sigma_{ij} = COV(X_i, X_j)$ e $\sigma_{ii} = COV(X_i, X_i) = Var(X_i) = \sigma_i^2$, si ha che:

$$Var(X) = \lambda^T \Sigma \lambda$$

vettore colonna che raggruppa i coefficienti λ_i

Limitandoci a combinare v.c. normali standard e indipendenti, in modi più complessi, si costruiscono altre distribuzioni utili nelle applicazioni:

- **DISTRIBUZIONE CHI-QUADRO**: Se Z_i (con $i = 1, \dots, n$) sono v.c. normali standard indipendenti, allora la v.c. X definita

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2 = \sum_{i=1}^{n} Z_i^2$$

è detta v.c. con distribuzione chi-quadro con n gradi di libertà (χ_n^2)

supporto $= \mathbb{R}^+$

Notiamo che il valore atteso del quadrato di una normale standard è:

$$\mathbb{E}[Z^2] = Var(Z) + \mathbb{E}^2[Z] = \sigma^2 + \mu^2 = 1^2 + 0^2 = 1 \text{ , quindi:}$$

→ Valore Atteso $\quad \mathbb{E}[X] = n \quad \left(\text{xk } \mathbb{E}[X] = \sum_{i=1}^{n} \overset{1}{\lambda_i} \mathbb{E}[Z^2] = n \cdot \overset{1}{\mathbb{E}[Z^2]} = n \right)$

→ Varianza $\quad Var(X) = 2n$

Nelle tabelle statistiche si trovano i quantili $\chi^2_{q,n}$ definiti dalla condizione

$$P\{X \leq \chi^2_{q,n}\} = q$$

La conseguenza pratica del teorema del limite centrale è che sommando un grande numero di contributi casuali, indipendenti tra di loro, si ottiene qualcosa che tende ad una distribuzione normale

Ad es, nel caso di tanti clienti che ordinano quantità indipendenti di un prodotto in un periodo di tempo, il teorema può giustificare l'assunzione di normalità per la domanda complessiva di prodotti di largo consumo

PROCESSI STOCASTICI: spesso si considera un insieme di v.c. che rappresentano successive realizzazioni nel tempo di uno stesso fenomeno. Si ha cioè a che fare con una collezione di variabili casuali indicizzate dal tempo

→ PROCESSO STOCASTICO A TEMPO DISCRETO: l'indice di tempo assume valori interi (come nel caso della domanda giornaliera di un prodotto)

X_t con $t = 0, 1, 2, 3, \dots$

In un certo senso un processo stocastico rappresenta una generalizzazione del concetto di funzione, in quanto esso assegna ad ogni valore della variabile indipendente t non un numero, bensì una v.c.

Osservando una possibile realizzazione di un processo stocastico, ovvero la realizzazione in sequenza di tutte le v.c., si ottiene una traiettoria campione (sample path)

Un aspetto fondamentale dei processi stocastici è che per descriverli non è sufficiente dare la distribuzione marginale delle v.c. $\forall t$, ma occorre anche considerare la distribuzione congiunta di tutte le v.c..
Per tale motivo nella pratica si considerano casi particolari e semplificati, come quello in cui le v.c. sono indipendenti tra loro:

PROCESSI GAUSSIANI: classe di processi stocastici molto comuni nelle applicazioni, il cui nome è dovuto al fatto che la distribuz. marginale delle diverse v.c. X_t che costituiscono il processo è normale

→ Questa classe contiene processi stocastici di diversa natura, ad es:

- Il processo stocastico $X_t = t \cdot \epsilon$ dove $t = 0, 1, 2, 3 \dots$ e $\epsilon \sim N(0, 1)$

  è gaussiano: $X_t \sim N(0, t^2)$

  tuttavia è un processo un po' degenere, in quanto legato alla realizzazione di una singola v.c. (se consideriamo valore di X_t in un solo istante di tempo, possiamo affermare di conoscere tutta la realizz. dell'intero processo stocastico)

PROCESSI STOCASTICI A TEMPO CONTINUO : in cui il parametro tempo può assumere valori nel continuo

$X(t)$ con $t \geq 0$

→ decisamente più complessi da trattare formalmente rispetto a quelli discreti

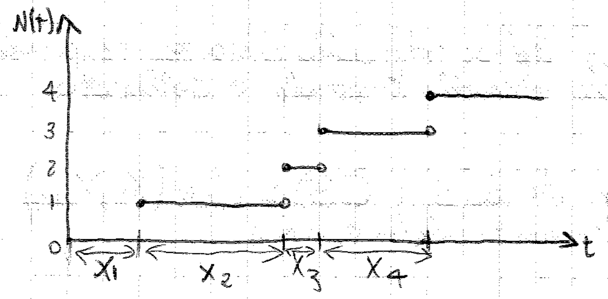
uno dei più comuni nelle applicazioni in logistica è il:

PROCESSO DI POISSON : ottenuto immaginando di osservare l'arrivo di clienti a un centro di servizio e di contare il n° di " arrivati fino all'istante t :

X_K (con $K = 1, 2, 3, 4 \dots$) → tempo di interarrivo tra il cliente $K-1$ ed il cliente K

→ Assunzione : le diverse X_t tutte indipendenti e distribuite esponenzialmente con tasso λ, che è il numero medio di clienti che arrivano nell'unità di tempo

Il processo stocastico $N(t)$ che conta il numero di clienti arrivati nell'intervallo $[0, t]$ è un processo a conteggio e presenta punti di discontinuità in corrispondenza dell'arrivo di un cliente



(una realizzazione del processo di Poisson)

Dato un interv. di tempo $[t_1, t_2]$ con $t_1 < t_2$, il n° di clienti che arrivano in tale intervallo, ovvero $N(t_2) - N(t_1)$, ha distribuz. di Poisson con parametro $\lambda(t_2 - t_1)$

Inoltre $N(t_2) - N(t_1)$ e $N(t_4) - N(t_3)$ sono indipendenti tra loro se $[t_3, t_4]$ e $[t_1, t_2]$ sono disgiunti

Il processo di Poisson è un modello utile x rappresentare fenomeni di arrivo di tanti clienti che non hanno relazione alcuna tra di loro

→ Qst è una conseguenza della proprietà di assenza di memoria della distribuzione esponenziale

Introducendo un tasso di arrivo $\lambda(t)$ variabile nel tempo si ottiene un PROCESSO DI POISSON NON OMOGENEO

Inoltre i diversi clienti, ad es. di un supermercato, possono chiedere qtà diverse di un prodotto

Se immaginiamo che ogni cliente in arrivo chieda un numero di pezzi descritto da una v.c. discreta, e consideriamo la domanda cumulata $D(t)$, ovvero tutti i pezzi richiesti nell'intervallo di tempo $[0, t]$, si ottiene il PROCESSO DI POISSON COMPOSTO

Poiché usiamo una v.c. per stimare il valore di un numero sconosciuto, è naturale chiedersi quali sono le:

PROPRIETÀ DI UN BUON STIMATORE:

- CORRETTO (o NON DEVIATO): il valore atteso dello stimatore dovrebbe coincidere con il valore sconosciuto (es. $\mathbb{E}[\hat{\mu}] = \mu$)
- ACCURATO → ERRORE QUADRATICO MEDIO $MSE_{\hat{\mu}}$ (o $See_{\hat{\mu}}$) $= \mathbb{E}[(\hat{\mu} - \mu)^2]$

Infatti la non deviatezza è solo UNA delle proprietà dello stimatore che, in quanto v.c., ha una sua varianza (che si vorrebbe fosse più piccola possibile)

↳ VARIANZA DELLA MEDIA CAMPIONARIA:

$$Var[\bar{X}] = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

dove σ^2 è la varianza del singolo elemento X_i del campione

OSS. La varianza dello stimatore scende al crescere della dimensione del campione

La varianza della media campionaria è utile per trarre indicazioni su quanto è affidabile la nostra stima

Il problema è che difficilmente la varianza σ^2 è nota, se non è neanche noto il valore atteso μ, per cui anche essa va stimata

Uno stimatore naturale della varianza è la

- VARIANZA CAMPIONARIA: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

↳ si tratta infatti di uno scarto quadratico medio risp. alla media campionaria

NB La varianza campionaria è una statistica, ovvero una v.c., mentre la varianza è un numero, eventualmente sconosciuto

OSS il denominatore $n-1$ è giustificato dal fatto che usiamo n dati per stimare la varianza, ma anche il valore atteso è stimato sulla base di questi dati, il che fa perdere un grado di libertà

Poiché gli stimatori sono v.c., è naturale chiedersi da quale distribuzione di probabilità essi siano caratterizzati

↳ una risp valida in generale NON esiste, infatti la distribuzione di una somma di v.c., seppure indipendenti e identicamente distribuite (i.i.d.), non ha una relazione banale con la distribuzione dei singoli addendi

Però il teorema del limite centrale ci dice che se la dimensione del campione è grande, la distribuz. della media campionaria tende ad una normale. Inoltre, se i singoli elementi del campione sono normali, Questa definizione implica anche $S_{XY} = S_{YX}$ e $S_X^2 = S_{XX}$

- **COEFFICIENTE DI CORRELAZIONE CAMPIONARIO tra X e Y:**

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \qquad -1 \leq r_{XY} \leq +1$$

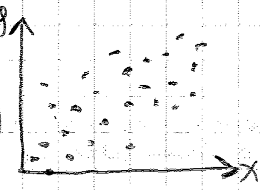
OSS. Ovviamente la dimensione minima del campione x poter stimare questi parametri è pari a due (altrimenti non sarebbe possibile calcolare alcuna misura di deviazione dalla media campionaria)

Il coeff. di correlazione è una misura adimensionale di quanto le variabili X e Y abbiano delle fluttuazioni in qualche misura "coordinate"

In altre parole, misura se ad uno scostamento positivo dell'osservazione di Y dalla sua media campionaria $\bar{Y}$ corrisponde uno scostamento positivo (o negativo) di X dalla sua media campionaria $\bar{X}$

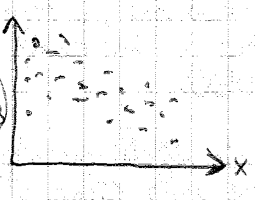
→ CORRELAZIONE POSITIVA tra Y e X: un coeff. di correlaz. positivo suggerisce che, in media, quando $Y > \bar{Y}$ anche $X > \bar{X}$; altrettanto " $Y < \bar{Y}$ " $X < \bar{X}$ in media

(es. $r_{XY} = 0.8$)



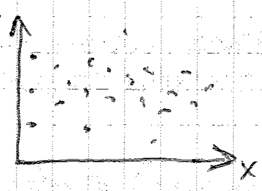
→ CORRELAZIONE NEGATIVA tra Y e X: un coeff. di correlaz. negativo, suggerisce che, in media, quando $Y > \bar{Y}$, si ha $X < \bar{X}$ e viceversa (" $Y < \bar{Y}$, " " $X > \bar{X}$)

(es. $r_{XY} = -0.8$)



→ CORRELAZIONE NULLA tra Y e X: quando la presenza di uno scostamento positivo tra Y e $\bar{Y}$ non comporta, in media, valori di X né superiori né inferiori alla sua media ($r_{XY} = 0$)

(es. $r_{XY} = 0$)



Per quanto riguarda la previsione, un coeff. di correlazione prossimo a zero significa che la variabile X non sembra essere una buona candidata x aiutare il previsore a comprendere quali saranno i valori futuri della domanda Y

Al contrario, la presenza di una correlazione forte (in valore assoluto) è segnale che le due variabili hanno delle oscillazioni intimamente legate e, quindi, la variabile X potrebbe essere molto utile nel prevedere la variabile Y.

dei campioni.

Se si ipotizza un campione normale, con valore atteso e varianza ignoti e stimati da $\bar{X}$ e S, sappiamo che

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Ne segue, indicando con $t_{1-\frac{\alpha}{2},n-1}$ un quantile della distribuzione t, che:

$$P\left\{-t_{1-\frac{\alpha}{2},n-1} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{1-\frac{\alpha}{2},n-1}\right\} = 1-\alpha$$

$$\Downarrow$$

$$P\left\{\bar{X} - t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right\} = 1-\alpha$$

$\Rightarrow \left(\bar{X} - t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}},\ \bar{X} + t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right)$ è un intervallo di confidenza per il valore atteso μ, con un livello di confidenza $1-\alpha$

Se ripetessimo l'estrazione del campione un numero grande di volte, il valore vero cadrebbe nell'intervallo di confidenza in una frazione pari a circa $(1-\alpha)\%$ dei casi.

Da un P.V. qualitativo si può osservare quanto segue:

- l'IF è tanto più grande quanto più grande è il livello di confidenza $1-\alpha$
- qndo il n° di campioni è molto grande, è possibile utilizzare i quantili $z_{1-\frac{\alpha}{2}}$ della distribuzione normale (usare, in qst caso, i quantili della t avrebbe carattere conservativo xk ne risulterebbe un IF più ampio risp. a qnto si avrebbe con i quantili z)
- l'IF è tanto più grande quanto più grande è la varianza σ dei singoli elementi del campione
- l'IF si riduce al crescere del numero n di campioni

OSS. Aumentare il numero di campioni non è l'unico modo per migliorare la qualità di una stima; infatti esistono:
- particolari procedure di campionamento (nel campo della simulazione stocastica su computer), dette metodi di riduzione della varianza, che permettono di ridurre la varianza σ^2 dei campioni
- stimatori alternativi caratterizzati da diverse varianze (a volte si preferisce utilizzare uno stimatore deviato a patto che la sua deviazione tenda a zero al crescere del numero di campioni e che la sua varianza sia inferiore a quella di stimatori alternativi non deviati)

NB Tutto quanto è stato detto circa gli IF vale <u>sotto ipotesi di indipendenza e normalità del campione</u>

Se l'ipotesi nulla è vera e se la popolazione è davvero normale, la statistica TS ha distribuzione t_{n-1}, quindi:

$$P_{\mu_0}\left\{-t_{1-\frac{\alpha}{2},n-1} \leq \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \leq t_{1-\frac{\alpha}{2},n-1}\right\} = 1-\alpha$$

probabilità assumendo H_0 vera

→ $IF\left(\bar{x}-t\cdot\frac{S}{\sqrt{n}},\ \bar{x}+t\cdot\frac{S}{\sqrt{n}}\right)$

→ Se il valore è davvero μ_0, la probabilità che il suo stimatore cada all'interno dell'IF è $1-\alpha$.

→ La regione critica del test, con livello di significatività α è proprio l'IF

Pertanto:

- Accetto H_0 se $\left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right| \leq t_{1-\frac{\alpha}{2},n-1}$

- Rifiuto H_0 " $\left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right| > t_{1-\frac{\alpha}{2},n-1}$

Il livello di significatività α ha una chiara interpretazione, ma può essere difficile individuarne un valore appropriato

Un concetto utile a questo scopo è il:

p-value: dato dalla probabilità che una variabile $T_{n-1} \sim t_{n-1}$ sia in valore assoluto maggiore della statistica TS:

$$p = 2\cdot P\{T_{n-1} \geq |TS|\}$$

→ moltiplico per 2 solo quando in H_0 c'è l'uguaglianza (es. $\mu = \mu_0$) in modo da considerare entrambe le code (se fosse $\mu \leq \mu_0$ o $\mu \geq \mu_0$ non moltiplicherei per 2)

→ fornisce un livello di significatività "critico" ($p = \alpha_{MAX}$), nel senso che l'ipotesi verrebbe accettata per tutti i valori $\alpha < p$

Ossia, solo correndo un rischio di errore pari ad almeno "p" possiamo rifiutare l'ipotesi

(b)

Ipotesi nulla $H_0: \mu \leq \mu_0$

" alternativa $H_A: \mu > \mu_0$

Il test si basa sulla stessa statistica ma considera una sola coda della distribuzione dello stimatore:

- Accetto H_0 se $\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \leq t_{1-\alpha,n-1}$

- Rifiuto H_0 " $\frac{\bar{X}-\mu_0}{S/\sqrt{n}} > t_{1-\alpha,n-1}$

Definiamo l'errore $e_i = y_i - f(x_i) = y_i - (a + bx_i)$

→ distanza verticale punto-retta

Nella scelta dei parametri a e b è necessario considerare una misura di errore aggregato, che cioè riassuma l'errore effettuato sulle diverse osservazioni i

Misuriamo l'errore totale come Scarto Quadratico:

$$SS = e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

→ usiamo il quadrato in modo che le distanze positive e negative non si elidano a vicenda

→ più semplice rispetto a misure in valore assoluto

misura l'accuratezza con cui il modello lineare rappresenta i dati empirici

Occorre quindi determinare a e b in modo tale da minimizzare questo errore

Per fare ciò, si impone che le derivate parziali di e rispetto ai due parametri si annullino:

$$\frac{\partial SS}{\partial a} = 2\left(\sum_{i=1}^{n}(y_i - a - bx_i)\right)\cdot(-1) = -2\left(\sum_{i=1}^{n}(y_i - a - bx_i)\right) =$$

$$= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bx_i = \sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow a = \frac{\sum_{i=1}^{n} y_i}{n} - b\frac{\sum_{i=1}^{n} x_i}{n} = \bar{y} - b\bar{x}$$

$$\frac{\partial SS}{\partial b} = 2\left(\sum_{i=1}^{n}(y_i - a - bx_i)\right)\cdot\left(-\sum_{i=1}^{n} x_i\right) = -2\sum_{i=1}^{n}((y_i - a - bx_i)x_i) =$$

$$= -2\sum_{i=1}^{n} x_i\left(y_i - (\bar{y} - b\bar{x}) - bx_i\right) =$$

$$= -2\sum_{i=1}^{n} x_i\left[(y_i - \bar{y}) - b(x_i - \bar{x})\right] = 0$$

$$\cancel{-2}\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \cancel{-2}\sum_{i=1}^{n} x_i b(x_i - \bar{x})$$

$$\Rightarrow b = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n} x_i(x_i - \bar{x})} =$$

→ 1° interpretazione

$$= \frac{\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \bar{y}}{\sum_{i=1}^{n} x_i \cdot x_i - \sum_{i=1}^{n} x_i \bar{x}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

→ 2° interpretazione

Si assume che il processo stocastico sia così strutturato:

$$Y_i = \alpha + \beta x_i + \varepsilon_i \qquad (\text{con } i = 1, \dots, n)$$

dove:

- ε_i è una v.c. di valore atteso nullo e st.dev σ_i
- v.c. ε_i sono i.i.d., ossia $\sigma_i = \sigma_\varepsilon \ \forall i$
- v.c. ε_i indipendenti da x_i (numeri)

> quindi $\varepsilon_i \sim N(0, \sigma_\varepsilon)$

↳ termine di RUMORE

↳ ossia conoscenza di x non è affetta da errori

Quindi:

Y_i è una v.c. tale che: $\mathbb{E}[Y_i] = \alpha + \beta x_i$, $Var(Y_i) = \sigma_\varepsilon^2$

In altre parole: si assume che il processo stocastico preveda una relazione lineare tra x e $\mathbb{E}[Y]$.
Tuttavia le singole osservazioni di Y non sono determinate esclusivamente da x ma anche da un fenomeno casuale ε_i che si caratterizza come un RUMORE (assunto stazionario ed indipendente sia da x che dalle altre estrazioni di ε) che si sovrappone alla relazione che lega x e $\mathbb{E}[Y]$.

Si assume, inoltre, che i parametri α e β del processo stocastico non siano noti (in caso contrario non avrebbe senso effettuare stime)

↳ Assumiamo solamente che processo è lineare e soddisfa le ipotesi sopra indicate

In queste condizioni possiamo conoscere il processo stocastico solamente attraverso le osservazioni (x, Y) che questo genera ed utilizzarle x effettuare delle stime circa i parametri che lo caratterizzano

Sotto queste ipotesi, cerchiamo ora di studiare le caratteristiche delle stime a e b ottenute minimizzando gli scarti quadratici

(N.B) Gli **STIMATORI a e b** sono a tutti gli effetti delle v.c. generate dal processo stocastico

↳ questo xk dipendono dalle singole osservazioni generate dal processo stocastico (le Y_i sono delle v.c.)

⇒ a e b (variabili casuali) sono delle stime dei parametri α e β (numeri incogniti)

Pertanto non è per nulla detto che la realizzazione a coincida con α e b con β

Valuteremo a e b in relazione alla loro capacità di fornire a chi non conosca i parametri del processo stocastico delle stime "buone" di α e β

Studiamo ora il problema dalla prospettiva di Dio in modo da saper poi valutare l'operato dell'uomo:

- Accuratezza (errori di stima)

Oltre a dimostrare che gli stimatori a e b non sono, in media, né ottimisti né pessimisti, è interessante verificare quanto siano accurati

In altri termini, può essere interessante analizzare la distribuzione delle v.c. $a-\alpha$ e $b-\beta$ non solo per il loro valore atteso (che abbiamo dimostrato essere nullo), ma anche per quanto riguarda la loro deviazione standard

Standard Error of Estimate (See): deviazione standard dell'errore di stima di un parametro

$$See_b \equiv \sqrt{Var(b-\beta)} = \sqrt{\mathbb{E}[(b-\beta)^2] - \underbrace{\mathbb{E}^2[b-\beta]}_{=0 \text{ (x la non deviatezza)}}} = \sqrt{\mathbb{E}[(b-\beta)^2]}$$

- Dal P.V. (ideale) di chi conosce il parametro oggetto della stima:

  la See è una misura della dispersione della stima intorno al parametro, cioè misura quanto la stima si avvicina o meno al parametro

- Dal P.V. (reale) di chi non conosce il "vero valore" del parametro oggetto della stima:

  la See fornisce delle indicazioni circa l'affidabilità della stima ottenuta e, quindi, permette di passare da una semplicistica visione di a come una stima puntuale, ad una più corretta interpretazione distributiva secondo la quale, in base ad un campione di osservazioni, non è in alcun modo possibile dire quale sia il parametro stimato, ma è solo possibile stimarne una distribuzione di probabilità

  tanto più la stima sarà buona e tanto più la distribuzione del parametro sarà "stretta" (cioè sarà molto poco probabile che si abbiano dei valori del parametro significativamente diversi dal valor medio della distribuzione stimata);
  al contrario, una stima "rozza" comporterà una distribuzione del parametro molto ampia (che suggerirà che il parametro "vero" potrebbe essere anche molto diverso dal valore atteso della distribuzione stimata)

  ⇒ se See_a (o See_b) è elevato, il vero valore di α (o β) potrebbe essere significativamente diverso da a(b);

  se " (") è piccolo, il parametro α (o β) è probabilmente assai prossimo ad a (o b)

Procediamo ora col quantificare gli errori di stima degli stimatori a e b:

$$E[(a-\alpha)^2] = var(a-\alpha) = var\left((\beta-b)\bar{x} + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right) =$$

$$= var((\beta-b)\bar{x}) + var\left(\frac{1}{n}\sum\varepsilon_i\right) + 2cov\left((\beta-b)\bar{x}, \frac{1}{n}\sum\varepsilon_i\right) =$$

$$= See_b^2 \cdot \bar{x}^2 + \frac{1}{n^2}\cdot n\sigma_\varepsilon^2 + 2cov\left((\beta-b)\bar{x}, \frac{1}{n}\sum\varepsilon_i\right) =$$

$$= \bar{x}^2 \cdot \frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2} + \frac{\sigma_\varepsilon^2}{n} + \frac{2\bar{x}}{n} cov\left(\beta-b, \sum\varepsilon_i\right) =$$

$$= \bar{x}^2 \frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2} + \frac{\sigma_\varepsilon^2}{n} + \frac{2\bar{x}}{n} cov\left(\frac{\sum(x_i-\bar{x})\varepsilon_i}{\sum(x_i-\bar{x})^2}, \sum\varepsilon_j\right) =$$

ho 1 v.c. a sx (ε_i) e 1 v.c. a dx (ε_j) x trovare se c'è cov tra le 2 sommatorie devo vedere se c'è cov tra le singole v.c. (ε_i e ε_j $\forall i \neq j$), ma x ip. sono i.i.d. x cui $cov(\varepsilon_i, \varepsilon_j) = 0$ $\forall i \neq j$, quindi ε_j posso scriverlo come ε_i e cov diventa var

$$= \quad '' \quad + \quad '' \quad + \frac{2\bar{x}}{n\sum(x_i-\bar{x})^2} cov\left(\sum(x_i-\bar{x})\varepsilon_i, \sum\varepsilon_i\right) =$$

$$= \quad '' \quad + \quad '' \quad + \frac{2\bar{x}}{n\sum(x_i-\bar{x})^2} \sum_{i=1}^{n} cov\left((x_i-\bar{x})\varepsilon_i, \varepsilon_i\right) =$$

$$= \quad '' \quad + \quad '' \quad + \frac{2\bar{x}}{n\sum(x_i-\bar{x})^2} \sum_{i=1}^{n}(x_i-\bar{x})var(\varepsilon_i) =$$

$$= \quad '' \quad + \quad '' \quad + \underbrace{\frac{2\bar{x}\,\sigma_\varepsilon^2}{n\sum(x_i-\bar{x})^2} \cdot \underbrace{\sum_{i=1}^{n}(x_i-\bar{x})}_{=0}}_{=0} =$$

x definizione di media (dato che $var(\varepsilon_i) = \sigma_\varepsilon^2$ non dipende dalle osservazioni x_i per ipotesi)

$$= \bar{x}^2 \frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2} + \frac{\sigma_\varepsilon^2}{n}$$

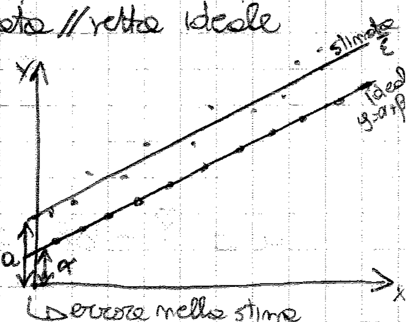
$$\Rightarrow \boxed{See_a = \sigma_\varepsilon \sqrt{\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2} + \frac{1}{n}}}$$

Nella formula si possono individuare 2 termini

1. $\sqrt{\frac{\sigma_\varepsilon^2 \bar{x}^2}{\sum(x_i-\bar{x})^2}}$ $= \bar{x}\cdot See_b$ legato agli errori dello stimatore b
2. $\sqrt{\frac{\sigma_\varepsilon^2}{n}} = \sqrt{var(\bar{\varepsilon})}$ legato al valore medio degli errori ε_i osservati

② Supponendo che $var(\beta-b) = 0$, ossia che $b = \beta$, cioè che la pendenza della retta sia corretta, avremo che retta stimata // retta ideale

sotto qsta ipotesi, l'unico motivo x il quale la retta stimata si discosta da quella ideale è dato dal termine $\frac{\sigma_\varepsilon^2}{n}$ → x cui al crescere di n, l'errore di posizionamento della retta diventerà meno probabile



errore nella stima dell'intercetta

- Se nel campione osservato gli errori sono stati tendenzialmente positivi ($\sum\varepsilon_i = \bar{\varepsilon} > 0$) tenderemo ad avere un valore $a > \alpha$ (quindi retta stimata sopra a retta ideale)
- Viceversa, se $\bar{\varepsilon} < 0$, risulterà uno stimatore $a < \alpha$ (e la retta di regressione si posizionerà al di sotto della retta ideale)

Pertanto calcolo lo scostamento quadratico medio tra i punti e la retta stimata $\hat{Y} = a + bx$ (dato che la retta ideale non è nota):

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - a - bX_i)^2}{n-2}}$$

OSS. Le n osservazioni costituiscono il set informativo con cui effettuiamo le stime dei nostri parametri (in termini statistici si usa il termine grado di libertà). La stima dei due parametri α e β "consuma" parte dell'informazione contenuta nelle n osservazioni; in particolare sono necessarie almeno 2 osservazioni (gradi di libertà) per effettuare queste stime. La stima di σ_ε avviene "successivamente" alla stima dei parametri α e β. Per questo motivo la stima di σ_ε non utilizza tutta l'informazione inizialmente disponibile ma solamente quella che è disponibile per questo scopo, cioè $n-2$ osservazioni (gradi di libertà).

Intuitivamente, se ho solo 2 punti ($n=2$) è evidente che la retta di regressione passerà esattamente per i 2 punti osservati. Ciò vuol dire che lo scostamento tra Y_i e $\hat{Y}_i$ sarà nullo, non perché la regressione non presenta variabilità, ma perché non abbiamo abbastanza info per stimarla! Potrò stimare la variabilità del processo solamente dal 3° punto in avanti, pertanto divido per $n-2$ (ossia perdo 2 g.d.l.)

INTERVALLI DI CONFIDENZA E TEST D'IPOTESI SUGLI STIMATORI

Per utilizzare al meglio le info ottenute attraverso la regressione, è necessario ora comprendere che tipo di distribuzione seguono le stime dei parametri

Per ipotesi le distribuzioni delle variabili ε_i che caratterizzano il nostro processo stocastico sono Normali, pertanto le stime a e b, derivando dalla somma di n variabili normali, mantengono questa caratteristica (P.V. DIO)

La conoscenza della distribuzione di probabilità degli stimatori viene utilizzata per:

- calcolare IF
- effettuare test d'ipotesi

Dobbiamo però tener presente che abbiamo a disposizione solo una stima della volatilità σ_ε, pertanto dovremo far riferimento alla distribuzione t di Student con $n-2$ g.d.l. (xk l'info residua per stimare σ_b è la stessa insita in σ_ε, ossia $n-2$) (P.V. uomo)

- Intervalli di Confidenza

Sebbene non siamo in grado di conoscere puntualmente il valore dei parametri α e β, utilizzando la distribuzione di probabilità oppure

$$P\left(-a - See_a \cdot t_{1-\frac{\alpha}{2}, n-2} \leq \alpha \leq -a + See_a \cdot t_{1-\frac{\alpha}{2}, n-2}\right) = 95\%$$

Stessa cosa per β, pertanto:

$$P(a_1 - t_{0.975,9} \cdot See_a \leq \alpha_1 \leq a_1 + t_{0.975,9} \cdot See_a) = 95\%$$

$$P(b_1 - t_{0.975,9} \cdot See_b \leq \beta_1 \leq b_1 + t_{0.975,9} \cdot See_b) = 95\%$$

. . . .

$\Rightarrow IF_{\alpha_1} = (99.642 - 2.26 \cdot 0.688,\ 99.642 + 2.26 \cdot 0.688) = (98.09,\ 101.19)$

$IF_{\beta_1} = (4.81,\ 5.33)$

$IF_{\alpha_2} = (55.19,\ 143.46)$

$IF_{\beta_2} = (-2.40,\ 12.52)$

OSS. A parità di probabilità p al crescere dello See, aumenta la dimensione dell'IF, cioè si ha una conoscenza meno precisa del parametro α o β

- Test di ipotesi

In base alle distribuzioni di probabilità ottenute è anche possibile effettuare dei test d'ipotesi sui parametri α e β

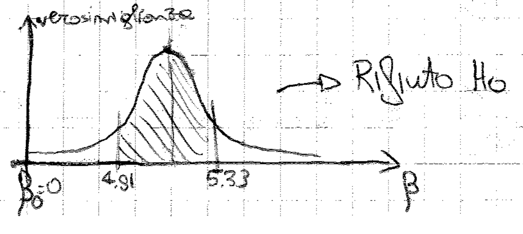
Le domande a cui possiamo fornire una risposta sono di 2 tipi diversi:

ⓐ Possiamo affermare con una probabilità pari a p che il parametro (α o β) goda delle proprietà oggetto delle nostre analisi (per es. >0 o <0 o $\neq 0$, a seconda dell'ipotesi da noi formulata?

- Nel caso in cui si voglia verificare se α o β sono diversi da zero (e quindi, nel secondo caso, se vi è un impatto statisticamente significativo di X su Y) con una probabilità pari a p, sostanzialmente dobbiamo verificare se nell'intervallo di confidenza, di probabilità p, si trova o meno il valore 0

Es. precedente Ipotesi nulla $H_0: \beta_1 = \beta_0 = \emptyset$ Ip. alternativa $H_1: \beta_1 \neq 0$

$\beta_0 \notin IF_{\beta_1}$ in quanto $\emptyset \notin (4.81,\ 5.33) \Rightarrow$ posso rifiutare H_0 con un livello di confidenza del 95%



Es. precedente) Nel 1° caso si trovano valori per α_1 e β_1 di

$$t_{p,q} = \frac{a_1}{See_{a_1}} = \frac{99.64}{0.688} = 144.84 \qquad t_{p,q} = \frac{b_1}{See_{b_1}} = \frac{5.07}{0.116} = 43.71$$

quindi si ha una probabilità sostanzialmente del 100% che questi due parametri siano $\neq 0$

difatti, $t_{p,q} = 43.71$ significa che affinché sia $b = 5.07$ dovrebbe verificarsi un evento che si discosti $43.71\,\sigma$ dalla media, il che ha probabilità pressoché nulla (difatti a qst distanze la massa probabilistica $\to \emptyset$)

Nel 2° caso troviamo

$$t_{p,q} = \frac{a_2}{See_{a_2}} = \frac{99.327}{19.53} = 5.08 \qquad t_{p,q} = \frac{b_2}{See_{b_2}} = \frac{5.063}{3.30} = 1.53$$

quindi $P\{\alpha_2 > 0\} \approx 100\%$, mentre $P\{\beta_2 > 0\} = 92\%$

perché si discosta solo di $1.53\,\sigma$ dal valor medio

In poche parole, col modus operandi (b), invece di cercare com'è fatto β, diamo per buono che $\beta = \beta_0$ e diciamo che deve essere:

$$P(\beta_1 - t_{n-2,1-\frac{\alpha}{2}} See_{b_1} \leq b_1 \leq \beta_1 + t_{n-2,1-\frac{\alpha}{2}} See_{b_1}) = 95\%$$

Se b non cade in questo range — o è solo sfortuna (ci troviamo nel 5%)
— o H_0 che abbiamo assunto vera è in realtà falsa

Difatti, col metodo (a) abbiamo rifiutato $H_0 : \beta = \emptyset$

dimostriamo che col metodo (b) arriviamo allo stesso risultato:

$$P_{\beta=0}(\emptyset - 2.26 \cdot 0.116 \leq b_1 \leq \emptyset + 2.26 \cdot 0.116) = 95\%$$

$$P(-0.262 \leq b_1 \leq 0.262) = 95\%$$

il nostro stimatore era $b_1 = 5.072$ pertanto al 95% H_0 è falsa!

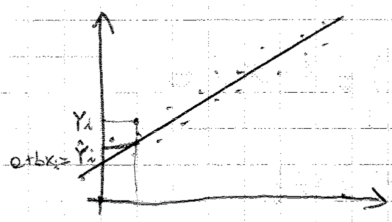
Stessa cosa per β_2: $\quad P(\emptyset - 2.26 \cdot 3.30 \leq b_2 \leq \emptyset + 2.26 \cdot 3.30) = 95\%$

$$P(-7.45 \leq b_2 \leq 7.45) = 95\%$$

noi abbiamo trovato $b_2 = 5.063 \implies b_2$ è compatibile con $H_0 : \beta_2 = \emptyset$

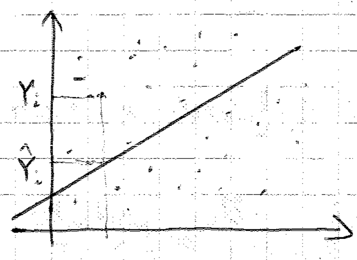
Se invece di domandarci se con una confidenza del 95% possiamo escludere $\beta_1 = 0$, ci chiediamo quale sia la max confidenza con la quale possiamo rifiutare $\beta_1 = \emptyset$ (e quindi affermare $\beta_1 \neq 0$):

STATISTICA R^2, ossia il quadrato del coefficiente di correlazione tra Y_i e $\hat{Y}_i$



Ossia, se le Y_i e le $\hat{Y}_i = a + bx_i$ (punti letti sulla retta) hanno fluttuazioni molto "coordinate" (si muovono insieme), allora la variabile esplicativa x è la variabile principale che spiega il fenomeno, ossia ho colto il grosso del fenomeno

↳ cioè se la mia stima letta sulla retta ($\hat{Y}_i$) si muove perfettamente insieme alle mie Y_i vuol dire che ho colto il fenomeno (o larga parte di esso)



Se, al contrario, si muovono l'uno indipendentemente dall'altro, vuol dire che c'è tanto del fenomeno che non ho ancora compreso

Qualora io abbia molti dati potrei avere a e b molto affidabili, ma ciò non vuol dire che io abbia colto larga parte del fenomeno

$$R^2 = r^2_{Y\hat{Y}} = \frac{\left[\sum_{i=1}^{n}(Y_i-\bar{Y})(\hat{Y}_i-\bar{\hat{Y}}_i)\right]^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2\cdot\sum_{i=1}^{n}(\hat{Y}_i-\bar{\hat{Y}}_i)^2}$$

Sapendo che $\bar{\hat{Y}} = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i = \frac{1}{n}\sum_{i=1}^{n}(a+bx_i) = \frac{\cancel{n}a}{\cancel{n}} + b\bar{x} = \bar{Y}$

è possibile riscrivere il coeff. di correlazione come:

$$R^2 = \frac{\left[\sum(Y_i-\bar{Y})(\hat{Y}-\bar{Y})\right]^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2} = \frac{\left[\sum_{i=1}^{n}\overbrace{[(Y_i-\hat{Y}_i)+(\hat{Y}_i-\bar{Y})]}^{\text{aggiungo e tolgo } \hat{Y}_i}(\hat{Y}_i-\bar{Y})\right]^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2} =$$

$$= \frac{\left[\sum_{i=1}^{n}(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y}) + \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})(\hat{Y}_i-\bar{Y})\right]^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}$$

possiamo ora dimostrare che il 1° termine a numeratore è NULLO:

$$\sum(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y}) = \sum_{i=1}^{n}[Y_i-(a+bX_i)](\cancel{a}+bX_i-\cancel{a}+b\bar{x}) =$$

$$= \sum_{i=1}^{n}[Y_i-(\bar{Y}-b\bar{x})-bX_i](bX_i-b\bar{x}) =$$

$$= b\sum_{i=1}^{n}[(Y_i-\bar{Y})-b(X_i-\bar{x})](X_i-\bar{x}) =$$

$$= b\left[\sum(Y_i-\bar{Y})(X_i-\bar{x}) - b\sum(X_i-\bar{x})^2\right] =$$

$$= b\left[(n-1)S_{XY} - \frac{S_{XY}}{\cancel{S_X^2}}(n-1)\cancel{S_X^2}\right] = \emptyset \quad \text{c.v.d.}$$

⇒ più $R^2 \to 1$ più il modello spiega il fenomeno

↳ variazioni di y (fenomeno) quasi completamente spiegate dalle X (variabile esplicativa)

↓

varianza non spiegata (quella dei punti intorno alla retta) $\to 0$

VERIFICA DELLE IPOTESI SUGLI STIMATORI

I risultati relativi alle proprietà degli stimatori sono validi solamente sotto le stringenti ipotesi fatte. Per questo motivo è importante verificare tali ipotesi per valutare se i risultati appena discussi si possono applicare ai dati a disposizione

↓

In base alle ipotesi del modello, il rumore presente nel modello ε_i è un processo di valore atteso nullo, stazionario, non autocorrelato e indipendente da X

↓

Per verificare in modo completo le ipotesi effettuate avremmo bisogno di conoscere le n estrazioni di ε_i, tuttavia, non conoscendo la retta $\alpha + \beta x$, è impossibile sapere quali siano state le effettive estrazioni di ε_i

↓

Per questo motivo siamo costretti ad usare una proxy di ε_i, cioè $e_i = y_i - (a + bx)$, ovvero gli scarti tra le osservazioni e la retta di regressione, detti RESIDUI

↓

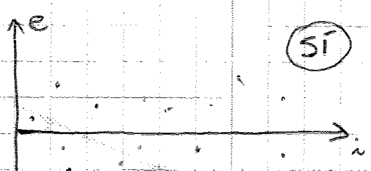
Se le ipotesi effettuate reggono, gli scarti e_i avranno valore atteso nullo, non dipenderanno da x e non saranno autocorrelati

↓

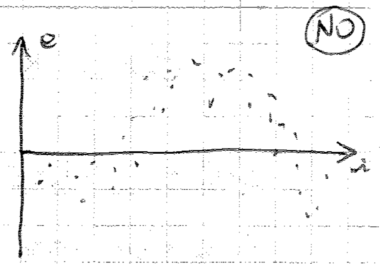
Per ciò che riguarda il valore atteso sappiamo che la stessa condizione $a = \bar{Y} - b\bar{x}$ garantisce che in media l'errore commesso sia pari a zero

Per ciò che riguarda invece le altre 2 ipotesi, si può procedere a valutare graficamente in primo luogo l'autocorrelazione e la stazionarietà del rumore:

- Se le ipotesi sono valide, la serie storica degli errori e_i dovrebbe avere un andamento simile a quello riportato in figura



- Affinché il fenomeno sia descrivibile con una retta di regressione, il fenomeno sottostante di dipendenza tra X e Y deve essere lineare; quindi i residui non devono essere influenzati dagli i come in figura e il fenomeno non deve presentare correlazione



A fronte di un set di dati che lascia supporre la presenza di una relazione funzionale non lineare tra Y e X è possibile effettuare delle trasformazioni in maniera tale da sfruttare la regressione lineare per stimare i parametri della supposta relazione tra Y e X

↓

è quindi, in primo luogo, necessario identificare una relazione funzionale tra Y e X che sia compatibile con i dati a disposizione e sia riconducibile al lineare (quindi caratterizzata solamente da 2 parametri)

Ad es, nel caso di relazione esponenziale del tipo:

$$Y = Kx^{\gamma}$$

si può passare alla relazione lineare

$$\log Y = \log(Kx^{\gamma}) = \log K + \gamma \log x$$

In altri termini, sarà possibile analizzare una relazione lineare tra Y e X sulla scala logaritmica

↓

la relazione tra log Y e log X appare essere effettivamente lineare e, quindi, è possibile pensare di utilizzare i dati trasformati per applicare il modello della regressione lineare

↓

Si cercherà quindi di stimare un modello lineare del tipo

$$\log Y = a + b\log X + e_i \quad \text{quindi} \quad a = \log K, \; b = \gamma$$

(NB) Utilizzando il modello della regressione lineare semplice sulle variabili trasformate (in qst caso logaritmiche), da un lato le ipotesi del modello dovranno essere verificate sulle stesse variabili trasformate (qui logaritmiche) e dall'altro, gli IF che saremo in grado di stimare saranno IF delle variabili trasformate e NON di quelle lineari; per ottenere gli IF delle variabili lineari sarà necessario effettuare le opportune trasformazioni

↓

inoltre è importante notare che nel modello lineare l'errore è additivo rispetto alla retta stimata, tuttavia, effettuando le trasformazioni inverse necessarie a riportare il modello in funzione di X e di Y, si avrà un errore NON più additivo, bensì avrà impatto su un fattore moltiplicativo 10^{e_i}

↓

perciò gli errori non saranno indipendenti da X, in quanto gli errori percentuali saranno stazionari e quindi gli errori tenderanno ad aumentare linearmente con X

- <u>Frequenza di aggiornamento</u> : indica ogni quanti t.b. aggiornare la previsione

→ se f = 1 si parla di <u>Rolling forecast</u>

↓

l'azienda prevede sempre per 52 sett (se horizon = 52), pertanto l'orizzonte si sposta in avanti nel tempo ad ogni aggiornamento (costo della previsione sostanzialmente maggiore)

→ se f = 52 si parla di <u>Fixed horizon</u>

↓

l'azienda prevede x 52 sett solo subito dopo aver generato la previsione, ma l'orizzonte temporale diminuisce progressivam. da 52 a 1 sett

OSS Ha senso aggiornare la previsione solo quando sono disponibili info aggiuntive

↓

la frequenza appropriata di aggiornamento dipende dal costo del processo di previsione, dalla disponibilità di info aggiuntive e dei benefici potenziali di una previsione più fresca

- <u>Prodotto</u> : per il quale facciamo la previsione

↓

più disaggrego la domanda più complico la previsione

- <u>Mercato</u> : mercato o area geografica alla quale facciamo riferimento

↓

più è ampio il mercato più è precisa la previsione

↓

prevedere a livello di singolo negozio è tutt'altro che banale xk la domanda è più bassa e tende ad essere più variabile. Inoltre, fattori esogeni come meteo locale o semplici lavori stradali possono cambiare significativamente l'andamento della domanda

↓

prevedendo x tutti i negozi della catena, invece, le oscillazioni si medieranno

Il processo previsionale è parte del più ampio processo decisionale: la previsione non è altro che uno strum x migliorare le decisioni, pertanto una " che non serve a prendere decisioni aziendali è inutile

↓

Per fissare correttamente i parametri del processo previsionale, bisogna prima capire il problema decisionale che vogliamo supportare con la nostra previsione

Il problema previsionale è impostato correttamente solo una volta che sono stati fissati correttamente il time bucket, l'orizzonte temporale, la frequenza

Se si usassero le vendite x prevedere la domanda, basse vendite ridurrebbero le previsioni che quindi guiderebbero i pianificatori a ridurre le scorte e scorte basse potrebbero portare ad una riduzione delle vendite

Quasi sempre domanda ≠ vendite, ad es. uno stockout di un dato prodotto può perturbare l'andamento delle vendite di altri prodotti dal momento che alcuni clienti, non trovando il prodotto desiderato, ne acquistano un surrogato (sostituto)

finirò col prevedere una domanda + alta x il surrogato quando invece il cliente voleva acquistare l'altro prodotto

③ Analisi della domanda: bisogna studiare e identificare l'andamento della domanda

tutte le tecniche quantitative di previsione fanno alcune assunzioni sul comportamento e andamento della domanda. Pertanto bisogna prima comprenderne il comportamento effettivo x poi cercare una tecnica di previsione che si adatti

④ Scelta e Parametrizzazione della tecnica di previsione: non basta sapere come funziona una tecnica x poterla usare, ma bisogna conoscere la logica che ne sta alla base x poterla adattare alle varie condizioni

Inoltre, l'efficacia di molti modelli dipende dalla scelta di valori adeguati dei parametri

il miglior modello x prevedere la domanda passata è anche " " " " " " " futura, a meno che non ci si aspetti un cambiamento significativo nella domanda

⑤ Previsione: scelta la tecnica e fissati i parametri, si genera la previsione

⑥ Misura della performance di previsione (misura dell'errore): si va a confrontare la previsione con la domanda che si è realmente verificata.

a) Se la previsione non è stata buona, torno alla fase ④

b) gli errori di previsione possono essere usati x giudicare la qualità del lavoro dei previsori e, tramite incentivi, portarli verso miglioramenti continui

c) incertezza della domanda σ_d = incapacità di prevedere la