

Lezioni di Topografia
Parte II - Il trattamento statistico delle misure

A. Manzano

*Dipartimento di Georisorse e Territorio
Politecnico di Torino, dicembre 2000*

DISPENSE DI TOPOGRAFIA

PARTE II – IL TRATTAMENTO STATISTICO DELLE MISURE

A. MANZINO

INDICE

PARTE SECONDA – IL TRATTAMENTO STATISTICO DELLE MISURE

6.	STATISTICA DI BASE.....	1
6.1	PRIMI TEOREMI DELLE DISTRIBUZIONI DI PROBABILITÀ.....	3
	a. Teorema della probabilità totale	3
	b. Definizione di probabilità condizionata.....	4
	c. Definizione di indipendenza stocastica.....	4
6.2	VARIABILI CASUALI	4
	Esempio di variabile casuale continua	5
	Funzione densità di probabilità	6
	Dalla variabile casuale alla variabile statistica.....	7
	La costruzione di istogrammi	8
	La media.....	9
	La varianza	10
6.3	TEOREMA DI TCHEBYCHEFF	11
	<i>Teorema</i>	11
	Il teorema nel caso di variabili statistiche.....	12
6.4	LA VARIABILE CASUALE FUNZIONE DI UNA VARIABILE CASUALE.....	13
	Esempio 1	15
	Esempio 2	16
6.5	TEOREMA DELLA MEDIA.....	16
	<i>Corollario 1</i>	16

	<i>Corollario 2</i>	17
	Esempio.....	18
6.6	LEGGE DI PROPAGAZIONE DELLA VARIANZA.....	18
	Osservazioni al teorema di propagazione della varianza.....	18
	Esempio di applicazione del teorema di propagazione della varianza..	19
6.7	ALCUNE IMPORTANTI VARIABILI CASUALI	19
	Distribuzione di Bernoulli o binomiale.....	19
	Distribuzione normale o di Gauss.....	21
	La distribuzione χ^2 (chi quadro).....	22
	Distribuzione t di Student	24
	La distribuzione F di Fisher	25
7.	LA VARIABILE CASUALE A n DIMENSIONI	27
	Esempio 1	28
	Esempio 2	29
7.1	DISTRIBUZIONI MARGINALI.....	30
7.2	DISTRIBUZIONI CONDIZIONATE.....	31
7.3	INDIPENDENZA STOCASTICA	32
	Leggi relative alle distribuzioni.....	32
7.4	VARIABILI CASUALI FUNZIONI DI ALTRE VARIABILI CASUALI	33
	Trasformazione di variabili	33
	Esempio di applicazione della trasformazione ad un caso lineare	34
7.5	MOMENTI DI VARIABILI n -DIMENSIONALI	36
	Teorema della media per variabili casuali n -dimensionali	37
	<i>Corollario 1</i>	37
	<i>Corollario 2</i>	37
	Momenti di ordine (n_1, n_2, \dots, n_k) di una variabile casuale n -dimensionale.....	37
	La propagazione della varianza nel caso lineare ad n -dimensioni	39
	Esercizio 1	40
	Esercizio 2	41
7.6	LA LEGGE DI PROPAGAZIONE DELLA VARIANZA NEL CASO DI FUNZIONI NON LINEARI	42
	Esercizio 3	43
	La propagazione della varianza da n dimensioni ad una dimensione .	45
	Esercizio 1	45
	Esercizio 2	45
	Esercizio 3	46
	Esercizio 4	46
	Esercizio 5	46
7.7	INDICE DI CORRELAZIONE LINEARE	47

7.8	PROPRIETÀ DELLE VARIABILI NORMALI AD n -DIMENSIONI	48
7.9	SUCCESSIONI DI VARIABILI CASUALI	52
7.10	CONVERGENZA «IN LEGGE»	53
7.11	TEOREMA CENTRALE DELLA STATISTICA	53
	Teorema	53
	Prima osservazione al teorema centrale della statistica	53
	Seconda osservazione al teorema centrale della statistica	54
7.12	LE STATISTICHE CAMPIONARIE E I CAMPIONI BERNOULLIANI	55
	Osservazione	55
	Definizione di <i>statistica campionaria</i>	55
7.13	LE STATISTICHE «CAMPIONARIE» COME «STIME» DELLE CORRISPONDENTI QUANTITÀ TEORICHE DELLE VARIABILI CASUALI	56
	Stima corretta o non deviata	56
	Stima consistente	56
	Stima efficiente	56
	Stima di massima verosimiglianza	56
7.14	FUNZIONE DI VEROSIMIGLIANZA E PRINCIPIO DI MASSIMA VEROSIMIGLIANZA	58
7.15	LA MEDIA PONDERATA (O PESATA)	60
8.	APPLICAZIONI DEL PRINCIPIO DEI MINIMI QUADRATI AL TRATTAMENTO DELLE OSSERVAZIONI	62
8.1	I MINIMI QUADRATI APPLICATI AD EQUAZIONI DI CONDIZIONE CON MODELLO LINEARE	64
	Esempio applicativo: anello di livellazione	65
8.2	MINIMI QUADRATI, FORMULE RISOLUTIVE NEL CASO DELL'UTILIZZO DI PARAMETRI AGGIUNTIVI	67
	Esempio applicativo	70
8.3	MINIMI QUADRATI : EQUAZIONI DI CONDIZIONE E PARAMETRI AGGIUNTIVI	72
8.4	PROPRIETÀ DELLE STIME \hat{y} ED \hat{x} , LORO DISPERSIONE	74
	Pure equazioni di condizione	75
	Pure equazioni parametriche	75
8.5	IL PRINCIPIO DEI MINIMI QUADRATI IN CASI NON LINEARI.....	76
8.6	ESERCIZIO	78
	Modello geometrico.....	79
	Modello stocastico e soluzione ai minimi quadrati.....	80

PARTE II – IL TRATTAMENTO STATISTICO DELLE MISURE

6. STATISTICA DI BASE¹

In questo capitolo ci doteremo di alcuni strumenti statistici per il trattamento delle misure.

Vediamo come si inserisce la statistica nella tecnica di misura e, per iniziare, come possiamo definire una misura. Conosciamo tre tipi di operazioni di misura:

- Misure dirette: vengono eseguite contando il numero di unità campione contenute in una quantità precostituita. Concettualmente funziona così ad esempio una bilancia a piatti, così è quando si misura col metro un oggetto ecc...
- Misure indirette: sono definite da un legame funzionale a misure dirette; ad esempio la misura indiretta della superficie del triangolo noti due lati e l'angolo compreso misurati direttamente. Il legame è nell'esempio $S = 1/2 ab \sin \gamma$.
- Misure dirette condizionate: sono delle misure dirette, ma fra loro sono legate da un legame funzionale interno. Ad esempio la misura diretta di tre angoli di un triangolo piano deve verificare la legge:

$$\alpha + \beta + \gamma = \pi$$

Nel capitolo 6 tratteremo prevalentemente le misure dirette, nel capitolo 7 quelle indirette (teorema della propagazione della varianza); infine le misure dirette condizionate saranno maggiormente trattate al capitolo 8 (minimi quadrati).

¹ Questa parte prende molti spunti, che liberamente interpreta, da «Fernando Sansò: Il trattamento statistico delle misure. - Clup 1990.» Da questo testo sono tratte inoltre dimostrazioni ed esempi.

L'operazione di misura, diretta o meno, ha in comune il fatto, che sotto opportune ipotesi, può essere considerata un'estrazione da una variabile casuale: vediamo infatti tre esempi che ci porteranno a giustificare questo paragone.

- a. Dato un corpo rigido di lunghezza poco maggiore di 3 m ed un metro campione suddiviso in mm, si desidera misurare il corpo con il metodo del riporto (o delle alzate).
- b. Il lancio di dadi non truccati.
- c. Si misurano le coordinate x, y del punto ove cade un proiettile su un bersaglio rettangolare sparato da uno stesso tiratore.

Questi esperimenti hanno in comune il fatto che, a priori, è impossibile predire in modo deterministico il risultato dell'esperimento: se si ripete infatti, si otterranno diversi risultati.

Nell'esempio a. il fatto che ripetendo l'operazione di misura si ottengano diversi risultati, porta a dire che in questa operazione si commettono degli «errori», negli altri casi il diverso risultato è dovuto alle variazioni non note dell'ambiente esterno e dell'oggetto di misura (e di come questi interagiscono), o ad una sua scarsa conoscenza globale e puntuale del fenomeno.

Questi «errori» possono classificarsi in:

- Errori grossolani: sono i più banali anche se spesso i più difficili a individuare. Possono essere ad esempio il mancato conteggio di una alzata, la trascrizione errata di una misura, la codifica errata di un punto, ecc.

I rimedi per evitarli sono l'acquisizione e il trattamento automatici, il controllo e la ripetizione delle misure possibilmente indipendenti ed ancora automatici. Non sono questi gli «errori» a cui intendiamo riferirci nell'esempio a.

- Errori sistematici: sono dovuti ad esempio all'imperfetta taratura dello strumento di misura o legati ad errori di modello (ad es. la misura indiretta di un angolo di un triangolo piano quando questo sia in realtà meglio «modellabile» sulla superficie ellissoidica), hanno la caratteristica di conservare valore e segno: nell'esempio a. la misura con più alzate tra due punti A e B, sarà sempre superiore alla reale, se i punti intermedi non sono esattamente sull'allineamento AB.

Sono eliminabili con tarature, con opportune procedure operative, o rendendoli di segno alterno (cioè pseudo accidentali): si può usare nel caso della bilancia non rettificata, ad esempio, il metodo della doppia pesata. Anche questi «errori» non sono quelli che giustificano i diversi risultati degli esperimenti a. b. e c.

- Fluttuazioni accidentali: sono a priori imprevedibili, sono di segno alterno e dipendono in senso lato dall'ambiente.

La fluttuazione accidentale della misura è un fenomeno *aleatorio* (casuale, probabilistico). Sono questi gli «errori» commessi negli esperimenti descritti. La scienza che studia questi fenomeni è la statistica matematica, perciò ne forniremo i concetti di base utili al trattamento delle misure geo-

detiche e topografiche. Ora cerchiamo di capire meglio in che ambito si cala la statistica nel trattamento delle misure. Potremmo definire la statistica la scienza che tenta di descrivere con certezza l'incertezza.

Nell'esempio del metro, notiamo che, se avessimo preteso di stimare la lunghezza del corpo al mm, avremmo ottenuto numeri apparentemente più variabili, mentre, chiedendo la misura al cm, il risultato sarebbe stato sempre uguale. Ne segue che, per la misura di una grandezza, l'indeterminazione si presenta solo con procedure di misura che spingono l'approssimazione ai confini delle capacità di misura dell'apparato usato.

Data per scontata questa indeterminazione, dobbiamo tuttavia dire che ci aspettiamo un risultato poco disperso, o meglio una gamma di possibili valori ed un ordine di priorità tra di essi.

Questa priorità, espressa come numero reale compreso tra zero e uno si chiama *probabilità*. Ne diamo ora la più usata definizione detta assiomatica che consiste nel definire la distribuzione di probabilità in base alle proprietà (assiomatiche) che deve soddisfare:

una distribuzione di probabilità P su un insieme S di valori argomentali, è una misura su una famiglia di sottoinsiemi di S (che include S stesso e l'insieme vuoto ϕ) che, oltre agli assiomi della misura:

$$\begin{cases} P(A) \geq 0 & 6.1 \\ P(\phi) = 0 & 6.2 \\ P(A \cup B) = P(A) + P(B) & 6.3 \end{cases}$$

soddisfa alla:

$$P(S) = 1 \quad 6.4$$

Vediamo un esempio pratico: il lancio della moneta. S è costituito da 2 valori argomentali che possiamo rendere numerici associando ad esempio $x = 0$ a «testa» ed $x = 1$ a «croce». S è l'insieme dei valori argomentali $\{0,1\}$ dei punti di coordinate $x=0$, $x=1$ sull'asse x .

I sottoinsiemi di S sono $\{\phi\}$, $\{0\}$, $\{1\}$, $\{0,1\}$.

Si ha $P(\{\phi\}) = 0$; $P(\{0\}) = 1/2$; $P(\{1\}) = 1/2$; $P(\{0,1\}) = 1$.

6.1 PRIMI TEOREMI DELLE DISTRIBUZIONI DI PROBABILITÀ

a. Teorema della probabilità totale

Dati due eventi A e B , sottoinsiemi disgiunti di S , la probabilità che si verifichi A o B , cioè $P(A \cup B)$ è:

$$P(A \cup B) = P(A) + P(B) \quad \text{se} \quad A \cap B = \phi \quad 6.5$$

Se A e B non sono disgiunti:

$$P(A \cup B) = P(A - B) + P(B) = P(A) + P(B) - P(AB) \quad 6.6$$

b. Definizione di probabilità condizionata

Si presenta quando si desidera esaminare la distribuzione solo su di una parte dei valori argomentali, restringendo S ad un sottoinsieme. Isolando una parte dei valori argomentali si genera un'altra distribuzione di probabilità.

Ad esempio in una popolazione di 100 persone caratterizzata dai possibili valori argomentali: capelli chiari o scuri, occhi chiari o scuri (vedi tabella 6.1), si desidera conoscere qual è la probabilità di estrarre una persona con occhi chiari fra quelle con i capelli chiari. Questa probabilità condizionata si indica $P(A|B)$ (probabilità di A condizionata a B) e vale:

$$P(A|B) = \frac{P(AB)}{P(B)} \tag{6.7}$$

Nell'esempio $P(B) = 50/100$, $P(AB) = 40/100$, $P(A|B) = 0.8$

Tab. 6.1

→	CAPELLI	C	S
Occhi	C	40	10
	S	10	40

c. Definizione di indipendenza stocastica

Diciamo A e B stocasticamente indipendenti se:

$$P(A|B) = P(A) \tag{6.8}$$

Per la 6.7 si ha:

$$P(A|B) = \frac{P(AB)}{P(B)} = P(A)$$

cioè:

$$P(AB) = P(A)P(B) \tag{6.9}$$

Dunque due eventi A e B sono stocasticamente indipendenti se e solo se la probabilità composta $P(AB)$ si scinde nel prodotto delle singole probabilità. Questa affermazione è il *teorema della probabilità composta*.

6.2 VARIABILI CASUALI

Definizione: una variabile casuale (vc) a una dimensione è una distribuzione di probabilità il cui insieme di valori argomentali S sia rappresentabile in \mathbf{R} , tale che sia definita la probabilità per qualunque insieme (ordinabile con x_0) del tipo:

$$I(x_0) = \{x \leq x_0\} \cap S \tag{6.10}$$

In questo modo sarà perciò caratterizzata dalla funzione di x_0 :

$$F(x_0) = P[x \in I(x_0)] \tag{6.11}$$

F prende il nome di funzione di distribuzione e gode delle proprietà:

$$F(x_0) \text{ è definita su } \forall x_0 \in \mathbf{R}$$

$$0 \leq F(x) \leq 1 \tag{6.12}$$

$$\lim_{x_0 \rightarrow -\infty} F(x) = 0; \quad \lim_{x_0 \rightarrow \infty} F(x) = 1 \tag{6.13}$$

$$F(x_2) \geq F(x_1) \quad \forall x_2 \geq x_1 \tag{6.14}$$

Una vc si dice discreta se l'insieme S è formato da un numero discreto di punti sui quali è concentrata una probabilità; se viceversa la probabilità che x assuma un *sin-golo* valore è sempre uguale a zero allora la vc è *continua*.

Nel primo caso avremo una funzione di distribuzione discontinua, nel secondo continua. Ad esempio il lancio di una moneta è rappresentato da una vc discreta:

i valori argomentali sono $x_1 = 0; x_2 = 1$; la variabile casuale x può rappresentarsi attraverso la tabella:

$$\left. \begin{array}{ll} x_1 = 0 & x_2 = 1 \\ p = 1/2 & p = 1/2 \end{array} \right\} \tag{6.15}$$

Per $x \leq 0$ $F(x) = 0$; per $0 < x \leq 1$ $F(x) = 1/2$ e per $x > 1$ $F(x) = 1$ e la sua funzione di distribuzione è disegnata in figura 6.1.



Fig. 6.1

Esempio di variabile casuale continua

Consideriamo una distribuzione di probabilità definita in $S = [0,1] \in \mathbf{R}$

$$P(a \leq x \leq b) = b - a = \text{cost} \tag{6.16}$$

Siamo nel caso di distribuzione uniforme, la sua funzione di distribuzione F, riportata in figura 6.2, sarà:

$$\begin{cases} F(x) = 0 & x \leq 0 \\ F(x) = x & 0 \leq x \leq 1 \\ F(x) = 1 & x > 1 \end{cases}$$

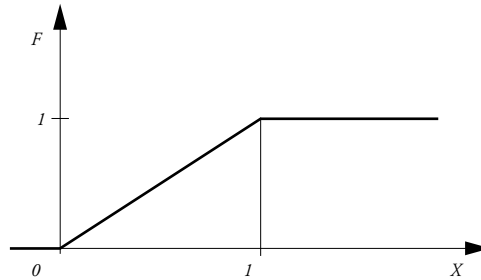


Fig. 6.2

Funzione densità di probabilità

Una qualunque variabile casuale può caratterizzarsi attraverso la sua funzione di distribuzione F . Se la vc è continua ci si chiede quale sarà la probabilità P che x sia compresa tra due valori $[x_0, x_0 + \Delta x]$. Si avrà:

$$P(x_0 \leq x \leq x_0 + \Delta x) = F(x_0 + \Delta x) - F(x_0) \quad 6.17$$

Se Δx è piccolo ed F differenziabile:

$$P(x_0 \leq x \leq x_0 + \Delta x) = dF(x_0) = F'(x_0)\Delta x = f(x_0)\Delta x$$

dove $f(x)$ vien detta *densità di probabilità* ed è funzione di x , si ha:

$$f(x_0) = F'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{P(x_0 \leq x \leq x_0 + \Delta x)}{\Delta x} \quad 6.18$$

che, per le caratteristiche di F , (monotona e crescente) sarà:

$$f(x_0) \geq 0 \quad \forall x$$

La funzione di distribuzione si ottiene allora come funzione integrale della densità di probabilità:

$$F(x) = \int_{-\infty}^x f(t) dt \quad 6.19$$

con l'ipotesi di normalizzazione (o standardizzazione, vedi 6.4):

$$\int_{-\infty}^{\infty} f(t) dt = 1 \quad 6.20$$

Si noti che:

$$\int_a^b f(x)dx = F(b) - F(a) = P(a \leq x \leq b)$$

Si abbia ad esempio la variabile casuale x definita così:

$$F = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

(vedi figura 6.2), la funzione densità di probabilità relativa è uniforme e vale:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x < 0; x > 1 \end{cases}$$



Fig. 6.3 – Funzione di densità di probabilità costante e uniforme.

Dalla variabile casuale alla variabile statistica

Se, per mezzo della variabile casuale si vuole rappresentare l'insieme dei possibili risultati di un esperimento non deterministico, si possono organizzare i dati in una tabella a doppia entrata in base ai risultati delle ripetizioni dell'esperimento.

Ad esempio:

$$\begin{cases} \text{testa} & \text{croce} \\ n_1 \text{ volte} & n_2 \text{ volte} \end{cases} \quad \text{con } n_1 + n_2 = N$$

Definiamo variabile statistica (vs) ad una dimensione la tabella di due sequenze di numeri che specifica come un dato si distribuisce fra la popolazione N :

$$\begin{cases} x_1 & x_2 \dots x_n \\ F_1 & F_2 \dots F_n \end{cases} \quad \text{ovvero} \quad \begin{cases} x_1 & x_2 \dots x_n \\ f_1 & f_2 \dots f_n \end{cases} \quad 6.21$$

x_i sono i valori argomentali, F_i le frequenze assolute ed $f_i = F_i/N$ le frequenze relative. Si ha:

$$\sum_1^n F_i = N; \quad \sum_1^n f_i = N \quad 6.22$$

Confrontando la 6.21 e la 6.22 si vede che la prima definisce una variabile casuale con distribuzione di probabilità concentrata sui valori $x_1 \dots x_n$, è sufficiente porre:

$$P(x = x_i) = f_i \quad 6.23$$

Con ciò, ogni definizione data e ogni proprietà mostrata per le variabili casuali deve valere anche per le variabili statistiche, poiché formalmente identificabili con le variabili casuali attraverso la 6.23.

La sostanziale differenza è di contenuto: sulla variabile casuale i numeri p_i associati ai valori x_i misurano un grado di possibilità che il risultato dell'esperimento abbia valore p_{ij} ; nel caso della variabile statistica il numero f_i registra a posteriori solamente il fatto che su N ripetizioni si sono ottenuti F_i risultati di valore x_i .

La probabilità, legata alla variabile casuale, è un ente aprioristico assiomatico, la frequenza, legata alla variabile statistica è un indice che misura a posteriori risultati empirici.

Per mezzo di questa identità formale, la funzione di distribuzione $F(x)$ delle variabili casuali, prende il nome, per le variabili statistiche, di funzione cumulativa di frequenza $F(x)$ e rappresenta la percentuale di elementi della popolazione il cui valore argomentale x_i risulta minore o uguale a x .

$$F(x) = \sum_i f_i = \frac{\sum N_i}{N} \quad \forall x_i \leq x \quad 6.24$$

La costruzione di istogrammi

Il concetto di densità di probabilità non è applicabile ad una variabile discreta perché la sua funzione di distribuzione è in ogni punto discontinua o costante.

Questo implica, per l'analogia tra variabili casuali e variabili statistiche che non si può definire un concetto analogo alla densità di probabilità per la variabile statistica.

È tuttavia importante poter confrontare la variabile statistica con particolari variabili casuali ben conosciute attraverso la funzione densità di probabilità, ciò si fa attraverso la costruzione di istogrammi.

Il confronto vien fatto tra probabilità (nella variabile casuale) e frequenza (della variabile statistica) in questo modo: si fissa un intervallo e si esamina la percentuale dei risultati che cadono nello stesso intervallo:

$$\Delta F(x_0) = \frac{N(x_0, \Delta x)}{N} \quad 6.25$$

dove il numeratore rappresenta il numero di elementi che cadono in detto intervallo. Il confronto è valido per N grande (ad esempio $N > 200$).

Si abbiano ad esempio una serie di valori nell'intervallo $I = (b-a)$.

Si riporta sull'asse x l'intervallo (a, b) e si divide in n parti (con $n < m$ valori dati), non necessariamente uguali (I_1, I_2, \dots, I_n) .

Per ogni intervallo si contano il numero di risultati che cadono in $I_i = N(I_i)$ e si sommano le frequenze relative a detto intervallo $\sum f_K = f_i$.

Si disegna sopra I_i un rettangolo di altezza $\sum f_K / I_i$.

Abbiamo costruito così una tabella:

$$\begin{cases} x_1 & x_2 \dots x_n \\ f_1 & f_2 \dots f_n \end{cases} \quad 6.26$$

dove x_i sono le ascisse dei valori medi degli intervalli I_i .

Si può verificare infine che:

$$\sum f_i = \sum_i \sum_K \frac{f_K}{I_i} I_i = 1 \quad 6.27$$

La media

La descrizione completa di una variabile casuale deriva dalla conoscenza della sua funzione di distribuzione o della densità di probabilità od altro di equivalente. Per molti usi pratici la vc è ben localizzata, cioè distribuita in una ristretta zona di valori ammissibili. Ad esempio, nella misura con distanziometri elettronici di distanze, una distanza di 1 km può avere ripetizioni che al più differiscono di 2-3 mm; per tutte queste variabili le informazioni più importanti da conoscere sono dove è localizzata la distribuzione e quanto è dispersa. Allo scopo, sono utili due indici: media e varianza.

Definizione: si chiama media della vc x , quando esista, il numero:

$$M[x] = \mu = \int_{-\infty}^{\infty} x f(x) dx \quad 6.28$$

Si noti l'analogia col momento statico di $f(x)$.

Nel caso di una vc discreta:

$$M[x] = \sum x_i p_i \quad 6.29$$

e, per analogia per una variabile statistica, la media, che si indica con m vale:

$$m = M[x] = \bar{x} = \sum x_i \quad f_i = \sum \frac{x_i N_i}{N} \quad 6.30$$

Dove con $M[\cdot]$ si intende l'operazione matematica (l'operatore) che, da una distribuzione, sia essa a priori vc o a posteriori vs , calcola un numero che è la media della distribuzione.

La 6.30 evidenzia in N_i il numero di volte che il valore argomentale x_i è stato estratto, presupponendo la costruzione di una tabella ordinata allo scopo, se invece con x_j indichiamo il singolo valore estratto si ha:

$$m = \bar{x} = \frac{1}{N} \sum x_i \quad j = 1, \dots, N \quad 6.31$$

Si può dimostrare che la media è un operatore lineare cioè gode delle proprietà:

$$M[x + y] = M[x] + M[y] \quad 6.32$$

$$M[kx] = k M[x] \quad 6.33$$

La varianza

È un indice che misura il grado di dispersione di una vc x attorno alla media.

Per definizione, se esiste vale

$$\sigma^2[x] = M[(x - \mu_x)^2] \quad 6.34$$

Si definisce la variabile scarto v

$$v = (x - \mu_x) \quad 6.35$$

La varianza si ottiene cioè applicando l'operatore media al quadrato della variabile scarto, in altri termini è il momento del secondo ordine della variabile scarto e si indica con $\sigma^2[x]$, σ_x^2 o solo σ^2 .

Per la variabile statistica, per analogia, la varianza si indica con $S^2(x)$, S_x^2 o solo S^2 . La radice quadrata della varianza si chiama scarto quadratico medio e si indica con sqm o con σ , tale valore è più usato della varianza, in quanto dimensionalmente omogeneo a x . Si ha dunque:

$$\sigma_x^2 = \int_{-\infty}^{\infty} (X - \mu_x)^2 f(X) dx \quad 6.36$$

e, per una vc discreta:

$$\sigma_x^2 = \sum_i (X_i - \mu_x)^2 p_i \quad 6.37$$

Con la solita analogia tra variabile casuale e variabile statistica, per quest'ultima si ha:

$$S^2 = \sum_i (X_i - M_x)^2 \frac{N_i}{N} = \frac{1}{N} \sum_j (X_j - M_x)^2 = \frac{\sum_j v_j^2}{N} \quad 6.38$$

Le ultime due espressioni valgono per una vc non ordinata: per questo si è sostituito l'indice j all'indice i .

Dalla definizione di varianza, tenendo conto della linearità dell'operatore media e sviluppando si ha:

$$\sigma_x^2 = M[X^2 - 2\mu X + \mu^2] = M[X^2] - 2\mu M[X] + \mu^2 = M[X^2] - \mu^2 \quad 6.39$$

che permette di calcolare σ^2 senza passare dalla variabile scarto. Per una vs non ordinata la 6.39 si trasforma:

$$S^2(X) = \frac{1}{N} \sum_j X_j^2 - m^2 \quad 6.40$$

Nella 6.39 rappresenta il momento del 2° ordine della vc che è dato dalla somma della varianza e del quadrato del valor medio.

6.3 TEOREMA DI TCHEBYCHEFF

Nell'analogia meccanica in cui la probabilità viene considerata come una distribuzione di massa concentrata o distribuita sull'asse x , la media esprime (a parte una costante di standardizzazione), la posizione del baricentro (il momento statico) e la varianza ha il senso di momento di inerzia rispetto al baricentro.

Più le masse sono disperse e più è alto il momento di inerzia, cioè la varianza. Questa nozione qualitativa è espressa in termini probabilistici quantitativi dal teorema di Tchebycheff che vale per qualsiasi tipo di distribuzione.

Teorema

Preso $\forall \lambda > 1$, e \forall variabile casuale x , vale la disuguaglianza:

$$P(|x - \mu_x| \leq \lambda \sigma_x) \geq 1 - \frac{1}{\lambda^2} \quad 6.41$$

Il teorema ci dice qual è la dimensione dell'intervallo $\lambda \sigma$ attorno alla media entro cui, per qualunque distribuzione di x , siamo sicuri di racchiudere una probabilità minima di $(1 - 1/\lambda^2)$.

Dimostrazione

Partiamo dalla definizione di σ_x^2 , cioè:

$$\sigma_x^2 = \sigma^2 = \int_{-\infty}^{\infty} (X - \mu_x)^2 f(x) dx$$

restringendo l'intervallo di integrazione sarà sempre vero che:

$$\sigma^2 \geq \int_{|x - \mu| \geq \lambda \sigma} (x - \mu)^2 f(x) dx \quad 6.42$$

Il primo termine all'interno dell'integrale varrà, per lo meno nell'intervallo di integrazione:

$$(x - \mu)^2 \geq (\lambda\sigma)^2$$

dunque l'espressione 6.42 varrà a maggior ragione sostituendo a $(x - \mu)^2$ la costante $(\lambda\sigma)^2$:

$$\sigma^2 \geq \lambda^2 \sigma^2$$

e, dividendo per σ^2 :

$$\frac{1}{\lambda^2} \geq \int_{|x-\mu| \geq \lambda\sigma} f(x) dx$$

cioè:

$$\frac{1}{\lambda^2} \geq P(|x - \mu_x| \geq \lambda\sigma)$$

c.v.d.

Il teorema nel caso di variabili statistiche

Consideriamo la variabile:

$$\begin{cases} x_1 \dots x_n \\ f_1 \dots f_n \end{cases}$$

e facciamo l'ipotesi che sia stata ordinata nel senso crescente

$$x_1 < x_2 \dots < x_n$$

per definizione:

$$s^2 = \sum_1^n (x_i - m)^2 f_i$$

Anche gli scarti v_i saranno allora crescenti. Possiamo dividere in tre parti la sommatoria di cui sopra:

$$s^2 = \sum_{i=1}^{v < \lambda s} v_i^2 f_i + \sum_{j=1}^{v \geq \lambda s} v_j^2 f_j + \sum_{k=1}^{\lambda - s < v < \lambda s} v_k^2 f_k$$

s^2 sarà sempre maggiore od uguale alle prime due sommatorie, cioè:

$$s^2 \geq \sum_{i,j/|v| \geq \lambda s} v_{i,j}^2 f_{i,j} \Rightarrow s^2 \geq \sum_{i,j/|v| \geq \lambda s} |v_{i,j}|^2 f_{i,j}$$

A maggior ragione, essendo nella sommatoria:

$$|v| \geq \lambda s, \text{ cioè } \lambda s < |v|$$

$$s^2 \geq \sum_{i,j/|v| \geq \lambda s} \lambda^2 s^2 f_{i,j}$$

dividendo entrambi i membri per s^2 :

$$1 \geq \lambda^2 \sum f_{i,j}$$

dividendo ancora per λ^2 e considerando che $\sum f_k = 1 - \sum f_{i,j}$:

$$\frac{1}{\lambda^2} \geq 1 - \sum f_k$$

cioè:

$$\sum f_k \geq 1 - \frac{1}{\lambda^2}$$

c.v.d.

6.4 LA VARIABILE CASUALE FUNZIONE DI UNA VARIABILE CASUALE

Seguiamo quest'esempio: sia x la vc che rappresenta il lancio di un dado non truccato, si ha, chiamando (p,d) i possibili eventi (pari o dispari):

$$P(x \in p) = \frac{1}{2}; \quad P(x \in d) = \frac{1}{2}$$

L'insieme S è costituito dall'unione di:

$$\{xp\} \cup \{xd\} = S$$

con

$$\{xp\} \cap \{xd\} = \phi$$

prendiamo ora una vc y che rappresenta il lancio di una moneta non truccata e leghiamola alla vc x con questa corrispondenza:

$$Y = g(X) = \begin{cases} x_p \leftrightarrow y \text{ testa} \\ x_d \leftrightarrow y \text{ croce} \end{cases}$$

essendo i possibili valori $1 \leq x_i \leq 6$ ed associamo per y_i valori numerici 0 e 1 a testa e croce.

Con ciò $0 \leq y_i \leq 1$. Si ha:

$$\begin{aligned} g(2) &= g(4) = g(6) = \text{testa} = 0 \\ g(1) &= g(3) = g(5) = \text{croce} = 1 \end{aligned}$$

Le due vc si esprimono allora:

$$X = \begin{cases} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{cases}$$

$$Y = \begin{cases} 0 & 1 \\ 1/2 & 1/2 \end{cases}$$

Questo esempio è stato fatto su variabili casuali discrete ma può generalizzarsi al caso di variabili continue in cui una funzione $y = g(x)$ sia definita su tutto l'insieme S_X dei valori argomentali della x .

La $g(x)$ trasforma lo spazio S_X nello spazio dei valori argomentali S_Y .

Cerchiamo ora invece una corrispondenza più interna, più puntuale: poniamo che la funzione $g(x)$ sia una funzione continua: quella tracciata ad esempio in figura 6.4.

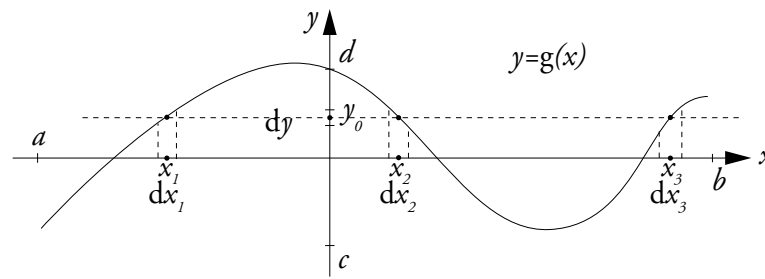


Fig. 6.4 – Variabile casuale funzione di variabile casuale.

dove il dominio dei valori argomentali è: $S_X = (a, b)$ $S_Y = (c, d)$.

Sia A_Y un sottoinsieme di S_Y ; a questo sottoinsieme corrisponderà un insieme:

$$A_X \in S_X / g(A_X) = A_Y$$

cioè, per definizione:

$$P(y \in A_Y) = P(x \in A_X) \tag{6.43}$$

Ed ora cerchiamo l'annunciata corrispondenza puntuale: scegliamo per A_Y un intervallo $dy(y_0)$ attorno a y_0 e, nell'ipotesi che $g(x)$ sia continua e differenziabile, si avrà che A_X sarà formata da uno o più intervalli attorno a x_i anch'essi di ampiezza dx_i , per cui si avrà la corrispondenza in termini probabilistici di:

$$A_Y = dy(y_0) \leftarrow \rightarrow A_X = \sum dx_i(x_i) \tag{6.44}$$

(con il simbolo Σ si intende qui l'operatore unione insiemistica \cup).

Si ha allora che:

$$P(y \in dy(y_0)) = \sum_1^m P(x \in dx_i(x_i))$$

cioè

$$f(y) dy = \sum f(x) |dx| \quad 6.45$$

in quanto per un intervallo infinitesimo il secondo membro è uguale a $f_X(x)|dx|$, dove $f_X(x)|dx|$ è la densità di probabilità della vc x . Dividendo entrambi i membri della 6.45 per $|dy|$ si ottiene:

$$\frac{P(y \in dy(y_0))}{|dy|} = \sum \frac{P(x \in dx_i(x_i))}{|dy|} = \sum \frac{f_X(x_i)}{\left| \frac{dy}{dx} \right|}$$

e, per definizione del primo membro:

$$f_y(y_0) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|} \quad 6.46$$

che è la formula di trasformazione di variabili casuali fra loro legate da una funzione g .

Esempio 1

Il legame fra due vc x ed y sia:

$$y = ax + b$$

si ha:

$$g'(x) = |a| ; \quad f_y(y) = \frac{f_X(x)}{|a|}$$

quel che serve tuttavia è avere una funzione esplicita di f_y in funzione di y cioè $f_y(y)$:

$$f_y(y) = \frac{f_X\left(\frac{y-b}{a}\right)}{|a|}$$

Se nell'esempio scegliamo per f_X la funzione definita normale standardizzata o Gaussiana:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad 6.47a$$

si avrà:

$$f_y(y) = \frac{1}{\sqrt{2\pi}|a|} e^{-\frac{1}{2}\left(\frac{y-b}{a}\right)^2} \quad 6.47b$$

Si può dimostrare che la media della vc y è b ed il suo sqm è $\pm a$. Attraverso la trasformazione lineare precedente si passa cioè dalla variabile non standardizzata alla variabile standardizzata di Gauss.

Esempio 2

Il legame sia $y = x^2$ cioè $x = \pm\sqrt{y}$. Ad un unico valore di y corrispondono due valori di x :

$$x_1 = -\sqrt{y}; \quad x_2 = \sqrt{y}$$

$$g'(x_1) = 2x_1 = -2\sqrt{y}; \quad g'(x_2) = 2x_2 = 2\sqrt{y};$$

$$f_y(y) = \sum \frac{f_x(-\sqrt{y})}{|-2\sqrt{y}|} + \frac{f_x(\sqrt{y})}{|2\sqrt{y}|} = \frac{f_x(-\sqrt{y}) + f_x(\sqrt{y})}{|2\sqrt{y}|}$$

Se, come sopra, $f_x(y)$ è la 6.47a si avrà:

$$f_y(y) = \frac{1}{2\sqrt{2\pi y}} e^{-\frac{1}{2}(-\sqrt{y})^2} + \frac{1}{2\sqrt{2\pi y}} e^{-\frac{1}{2}(\sqrt{y})^2} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} \quad (\text{per } y \geq 0) \quad 6.48$$

Il quadrato di una variabile gaussiana 6.47 ha dunque funzione di distribuzione di equazione 6.48 che vedremo essere la variabile χ^2 ad una dimensione cioè χ_1^2 .

6.5 TEOREMA DELLA MEDIA

Siano x ed y due variabili casuali legate dalla relazione $y = g(x)$; allora la media di y , se esiste vale:

$$\mu_y = M_y [y] = M_x [g(x)] \quad 6.49$$

È cioè possibile fare il cambiamento di variabili nell'operatore media $M[\cdot]$.

Dimostrazione

Poniamoci, solo per semplicità, nel caso che $g(x)$ sia monotona e crescente ($g'(x) > 0$). Ricordando la definizione di media e la 6.46:

$$M_y[y] = \int_{-\infty}^{\infty} y f_y(y) dy = \int_{-\infty}^{\infty} y \frac{f_x(x)}{g'(x)} dy = \int_{-\infty}^{\infty} g(x) \frac{f_x(x)}{g'(x)} \cdot g'(x) dx$$

$$M_y[y] = \int_{-\infty}^{\infty} g(x) f_x(x) dx = M_x[g(x)] \quad \text{c.v.d.}$$

Seguono due importantissimi corollari del teorema.

Corollario 1

La media è un operatore lineare, vale a dire se x ed y sono due vc ed

$$y = ax + b \Rightarrow M_y[y] = aM_x[x] + b \quad 6.50$$

Infatti:

$$M_y[y] = M_x[ax + b] = \int_{-\infty}^{\infty} (ax + b) f_x(x) dx = a \int_{-\infty}^{\infty} x f_x(x) dx + b \int_{-\infty}^{\infty} f_x(x) dx$$

$$M_y[y] = aM_x[x] + b$$

Corollario 2

Sia $y = g(x)$; sotto opportune ipotesi della g rispetto alle distribuzioni di x ed y e con una certa approssimazione vale:

$$\mu_y = M_y[y] = g(\mu_x) \tag{6.51}$$

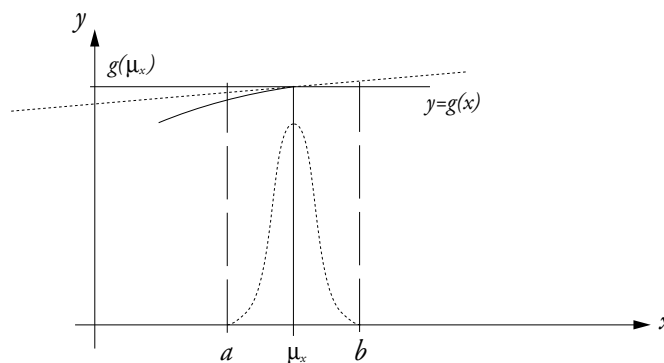


Fig. 6.5 – Dimostrazione del 2° corollario del Teorema della media.

Dimostrazione del 2° corollario

Sia x una vc abbastanza concentrata attorno a μ_x (che abbia cioè piccolo σ_x), supponiamo poi che $g(x)$ abbia andamento molto regolare attorno a μ_x , per lo meno in un intorno $[a, b]$.

Sviluppando $g(x)$ si ha, al primo ordine:

$$g(x) \cong g(\mu_x) + g'(\mu_x)(x - \mu_x)$$

$$\mu_y = M_y[y] = \int_{-\infty}^{\infty} g(x) f_x(x) dx \cong \int_{-\infty}^{\infty} [g(\mu_x) + g'(\mu_x)(x - \mu_x)] f_x(x) dx$$

$$g(\mu_x) \int_{-\infty}^{\infty} f_x(x) dx + g'(\mu_x) \int_{-\infty}^{\infty} (x - \mu_x) f_x(x) dx$$

Il secondo termine del secondo membro è nullo in quanto rappresenta la media della variabile scarto, risulta dunque provata la 6.51.

L'equazione 6.51 si trasforma nella 6.50 nel caso lineare, nel quale è rigorosa.

Esempio

Di un anello si è più volte misurato direttamente il diametro, ottenendo il valore medio di \bar{x} ; si desidera conoscere la superficie interna media in modo indiretto. Applicando la 6.51 si ha:

$$\bar{y} = \frac{\pi \bar{x}^2}{4}$$

6.6 LEGGE DI PROPAGAZIONE DELLA VARIANZA

Sotto le ipotesi del secondo corollario del teorema della media se la vc y è una funzione della vc x :

$$y = g(x)$$

$$\sigma_y^2 = g'(\mu_x)^2 \sigma_x^2 \quad 6.52$$

Dimostrazione

Poniamoci nel solito intervallo $[a, b]$ che comprende quasi tutto l'insieme S_X , nel quale valgono la 6.50 e la 6.51. Per funzioni monotone si ha $f_x(x)dx = f_y(y)dy$, dunque:

$$\sigma_y^2 = \int_a^b (g(x) - \mu_x)^2 f_x(x) dx$$

e, sviluppando $g(x)$:

$$\sigma_y^2 \cong \int_a^b [g(\mu_x) + g'(\mu_x)(x - \mu_x) - \mu_y]^2 f_x(x) dx$$

$$\sigma_y^2 \cong \int_a^b g'(\mu_x)^2 (x - \mu_x)^2 f_x(x) dx \cong g'(\mu_x)^2 \int_a^b (x - \mu_x)^2 f_x(x) dx$$

cioè a dire la 6.52.

Osservazioni al teorema di propagazione della varianza

La 6.52 è una formula rigorosa nel caso che $g(x)$ sia una funzione lineare; in tal caso infatti:

$$y = ax + b \Rightarrow \mu_y = a\mu_x + b$$

$$\sigma_y^2 = M_y[(y - \mu_y)^2] = M[(ax + b - a\mu_x - b)^2] = a^2 M[(x - \mu_x)^2]$$

$$\sigma_y^2 = a^2 \sigma_x^2 \quad \text{c.v.d.}$$

Data una variabile casuale x qualunque è sempre possibile con una trasformazione lineare costruire da questa una variabile casuale z tale che:

$$\mu_z = 0; \quad \sigma_z^2 = 1 \quad 6.53$$

detta variabile casuale standardizzata.

Grazie al teorema della media e della propagazione della varianza basta infatti porre:

$$z = \frac{x - \mu_x}{\sigma_x} \quad 6.54$$

e si avrà;

$$M[z] = \frac{1}{\sigma_x} M[x - \mu_x] = 0$$

$$\sigma^2[z] = \frac{1}{\sigma_x^2} \sigma_x^2 = 1$$

Esempio di applicazione del teorema di propagazione della varianza

Nel calcolo della superficie interna di un anello si è misurato il diametro medio $x = 5$ cm e stimato $\sigma_x = \pm 0.01$ cm si desidera calcolare la superficie media e la relativa varianza:

$$\bar{y} = \frac{\pi \bar{x}^2}{4} 19.63495 \text{ cm}^2$$

Quante cifre hanno senso in questo calcolo?

$$\sigma_y^2 = \left(2 \bar{x} \frac{\pi}{4}\right)^2 \sigma_x^2; \quad \sigma_y = \frac{\bar{x} \pi}{2} \sigma_x$$

$$\sigma_y = \pm 0.0785 \text{ cm}^2$$

Ha senso definire dunque \bar{y} al massimo a due cifre dopo la virgola:

$$\bar{y} = 19.63 \text{ cm}^2 \pm 0.078 \text{ cm}^2.$$

6.7 ALCUNE IMPORTANTI VARIABILI CASUALI

Distribuzione di Bernoulli o binomiale

Consideriamo un esperimento stocastico ε e siano S i suoi possibili risultati. Supponiamo che S sia costituita da due insiemi disgiunti A e B di eventi incompatibili 0 ed 1 aventi rispettivamente probabilità p e $q=(1-p)$:

$$P(A) = p; \quad P(B) = q; \quad \varepsilon := \begin{cases} 0 & 1 \\ q & p \end{cases} \quad 6.55$$

con:

$$M|\varepsilon| = \sum x_i p_i = p; \quad \sigma^2|\varepsilon| = (1-p)^2 p + (0-p)^2 q = pq$$

Da questa vc discreta ne costruiamo una seconda: consideriamo n ripetizioni indipendenti di ε ed indichiamo con β la vc discreta (intera) che descrive la probabilità che, su n esperimenti ε , k abbiano un risultato in A e $(n-k)$ un risultato in B. Per costruire la seconda riga della vc k :

$$\beta := \begin{cases} \beta = & 0 & 1 & 2 & 3 & \dots & n \\ & - & - & - & - & - & - \end{cases}$$

abbiamo ora bisogno di conoscere il teorema delle probabilità totali che dice in questo caso: la probabilità di k successi su n prove è uguale alla somma delle probabilità di $(k-1)$ successi su $(n-1)$ prove per la probabilità p di un nuovo successo, più la probabilità di k successi in $(n-1)$ prove per la probabilità q di un insuccesso.

È possibile cioè ricavare la formula ricorsiva:

$$P(n, k) = p \cdot P(n-1, k-1) + q \cdot P(n-1, k) \tag{6.56}$$

Partiamo da una prova dell'esperimento: la probabilità di successo sarà p e di insuccesso q :

$$P(1, 1) = p; \quad P(1, 0) = q \tag{6.57}$$

Si ha ad esempio, applicando la 6.56:

$$P(2, 0) = p \cdot 0 + q \cdot q = q^2$$

ed in genere $P(n, 0) = q^n$. Viceversa:

$$P(2, 1) = p \cdot P(1, 0) + q \cdot P(1, 1) = P(2, 1) = pq + pq = 2pq$$

$$P(1, 2) = p \cdot P(0, 1) + q \cdot P(0, 2) = 0$$

$$P(2, 2) = p \cdot P(1, 1) + q \cdot P(1, 2) = p^2$$

in genere $P(n, n) = p^n$ e, per valori qualunque di (n, k) si dimostra che vale:

$$P(n, k) = \binom{n}{k} p^k q^{n-k} \tag{6.58}$$

Dunque la vc discreta β è così definita:

$$\beta := \begin{cases} k = & 0 & 1 & 2 & \dots & n \\ & q^n & \binom{n}{2} p^2 q^{n-2} & \dots & \dots & p^n \end{cases} \tag{6.59}$$

Per ricavare media e varianza della 6.59 possiamo con maggior facilità applicare il teorema della media e quello della propagazione della varianza. Essendo β la somma delle n variabili ε :

$$\beta = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n$$

ed avendo ciascuna variabile ε media uguale a p e varianza uguale a pq :

$$M[\beta] = np \tag{6.60}$$

$$\sigma^2[\beta] = npq \tag{6.61}$$

La distribuzione binomiale ha la forma di figura 6.6 (è discreta e dunque costituita da un insieme distinto di punti).

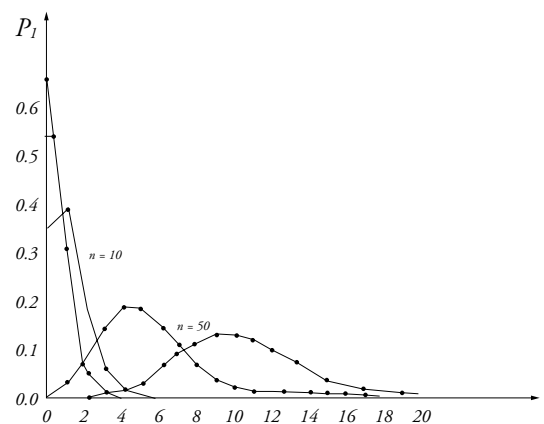


Fig. 6.6 – Distribuzione binomiale o di Bernoulli.

Distribuzione normale o di Gauss

La funzione densità di probabilità è data dalla:

$$f_x(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)^2} \quad -\infty \leq x \leq \infty \tag{6.62}$$

dove si può verificare che μ e σ^2 sono media e varianza della variabile casuale già vista nella 6.47. La figura 6.7 mostra due distribuzioni normali con stessa media, $\mu = 1$ ma con $\sigma = \pm 0.8$ e $\sigma = \pm 2$ rispettivamente.

La standardizzazione della 6.62 conduce alla variabile z con distribuzione:

$$f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{6.63}$$

Se cerchiamo la funzione di distribuzione della 6.63 si ha:

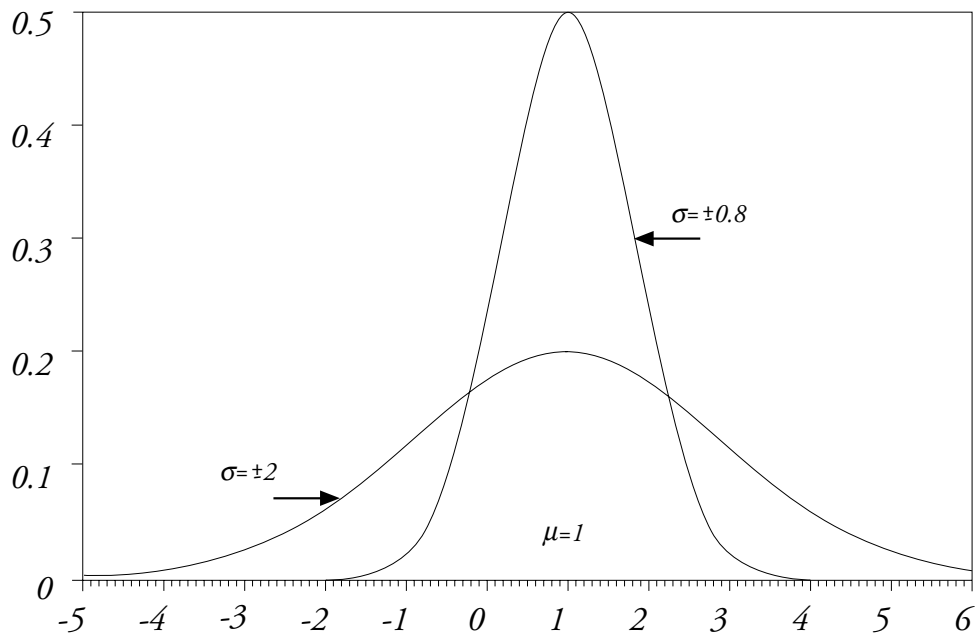


Fig. 6.7 – Distribuzione normale o di Gauss

$$\phi(z) = (\text{def})\text{erf}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz \tag{6.64}$$

Attraverso la $\Phi(z)$ possiamo ricavare la probabilità che z od x appartengano a vari intervalli attorno a σ : i valori più comuni sono:

$$\left. \begin{aligned} P(|x - \mu_x| < \sigma) &= \text{erf}(1) - \text{erf}(-1) = 0.683 \\ P(|x - \mu_x| < 2\sigma) &= \text{erf}(2) - \text{erf}(-2) = 0.954 \\ P(|x - \mu_x| < 3\sigma) &= \text{erf}(3) - \text{erf}(-3) = 0.997 \end{aligned} \right\} \tag{6.65}$$

La distribuzione χ^2 (chi quadro)

Si può dimostrare che se z_1, z_2, \dots, z_n sono n variabili casuali indipendenti, aventi una distribuzione normale e standardizzata la somma χ^2 dei loro quadrati è pure una variabile casuale:

$$\chi^2 = z_1^2 + z_2^2 + \dots + z_n^2 \tag{6.66}$$

la cui densità di probabilità (chiamando per non generare confusioni $\chi^2 \equiv h$) è data da:

$$f(h) = h^{(n/2-1)} e^{-h/2} [2^{n/2} \Gamma_{n/2}]^{-1} \tag{6.67}$$

Come si vede χ^2 dipende anche dal parametro intero n , detto grado di libertà. Nella 6.67 il termine entro la quadra è una costante che fa sì che la relativa funzione di distribuzione valga $\lim_{b \rightarrow \infty} F(b) = 1$.

Nelle 6.67, in parentesi, compare la funzione Γ di Eulero, generalizzazione della funzione fattoriale; per numeri reali si calcola attraverso:

$$\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx \tag{6.68}$$

Per valori di s semi-interi si usa la più comoda formula ricorsiva

$$\Gamma(1) = 1; \quad \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2} \tag{6.69}$$

$$\Gamma(p + 1) = p\Gamma(p) \tag{6.70}$$

Si dimostra che:

$$\mu(\chi^2) = n$$

$$\sigma^2(\chi^2) = 2n$$

Nella pratica occorre trovare la probabilità totale dei valori argomentali che superino χ_0^2 (figura 6.8).

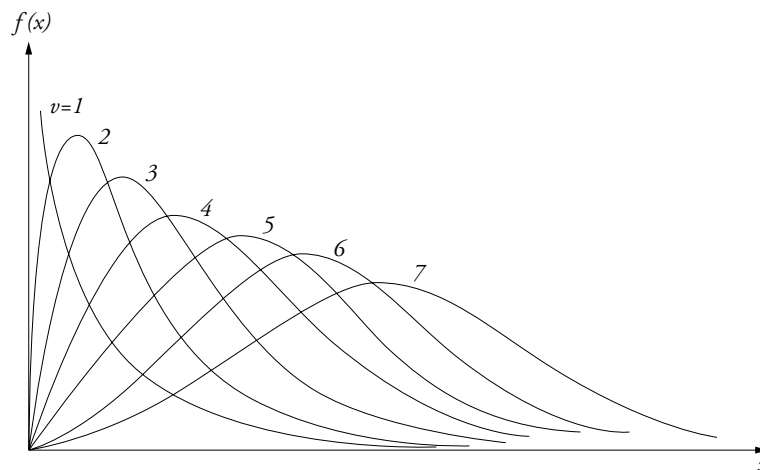


Fig. 6.8 – Funzione densità di probabilità χ^2 .

$$P(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} f(b) db \tag{6.71}$$

Questi valori sono in genere tabulati in funzione di χ_0^2 e di n . Tale variabile si indica spesso anche con χ_n^2 per evidenziare il numero di gradi di libertà.

Distribuzione t di Student

Sia z una normale standardizzata e z_i altre variabili normali standardizzate $i = 1 \dots n$ e sia:

$$y = \sqrt{z_1^2 + z_2^2 + \dots + z_n^2} = \sqrt{\chi_n^2} \quad 6.72$$

una seconda variabile casuale così costruita ed indipendente da z .

Si definisce la variabile t come:

$$t = t_n = \frac{z\sqrt{n}}{\sqrt{\chi_n^2}} = \frac{z\sqrt{n}}{\sqrt{z_1^2 + z_2^2 + \dots + z_n^2}} \quad 6.73$$

Si dimostra che la funzione densità di probabilità $f(t)$ vale:

$$f(t) = \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \left[\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma(n/2)} \right] \quad 6.74$$

La 6.74 è simmetrica rispetto all'origine, dunque:

$$\mu(t) = 0 \quad 6.75$$

Si prova che:

$$\sigma^2(t) = \frac{n}{n-2} \quad \text{per } n > 2 \quad 6.76$$

Per grandi valori di n , t è molto simile alla variabile z .

Per un certo valore del grado di libertà n i valori della funzione di distribuzione di questa variabile casuale si trovano tabulati in funzione delle probabilità $\alpha, t_{1-\alpha}^n$; ad esempio per $\alpha = 5\%$ si trova tabulato:

$$P(t < t_\alpha) = 1 - \alpha \quad 6.77$$

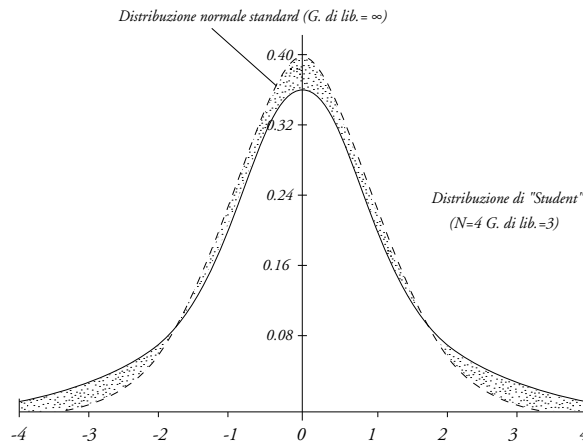


Fig. 6.9 – Distribuzione t di Student.

La distribuzione F di Fisher

Siano date due vc χ^2 ad n ed m gradi di libertà ed indipendenti tra loro; allora il rapporto

$$F = \frac{\chi_n^2 \cdot \frac{1}{n}}{\chi_m^2 \cdot \frac{1}{m}} = F_{n,m} \tag{6.78}$$

è una vc detta F di Fisher ad (n, m) gradi di libertà.

Si può dimostrare che:

$$f(F) = \frac{F^{(n-2)/2}}{(nF + m)^{(n+m)/2}} \cdot \frac{n^{n/2} m^{m/2} \Gamma\left(\frac{n+m}{2}\right)}{\Gamma(n/2) \Gamma(m/2)} \quad \text{per } F \geq 0 \tag{6.79}$$

e che:

$$M[F] = \frac{n}{n-2} \quad (n > 2) \tag{6.80}$$

$$\sigma^2(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \tag{6.81}$$

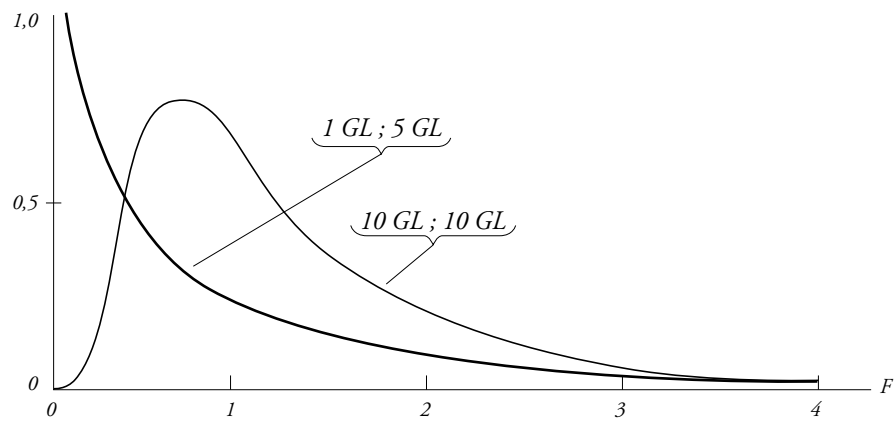


Fig. 6.10 – Variabile F di Fisher.

Anche qui le tabelle riportano $P(F \geq F_0)$ per (n, m) gradi di libertà.

Generalmente è impiegata la variabile \bar{F} detta di Fisher modificata che risulta essere sempre maggiore di 1 essendo così definita:

$$\bar{F} = \begin{cases} F & \text{con } F \geq 1 \\ 1/F & \text{con } F < 1 \end{cases} \quad 6.82$$

7. LA VARIABILE CASUALE A n DIMENSIONI

Partiamo col definire una variabile casuale discreta a n dimensioni cioè quella variabile per cui ogni valore argomentale può essere indicato come un vettore $x \in \mathbf{R}^n$, cioè un punto nello spazio \mathbf{R}^n :

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \tag{7.1}$$

L'insieme dei valori argomentali S sarà dunque un insieme $S \in \mathbf{R}^n$ in cui è definita la nostra distribuzione di probabilità.

La vc si dice discreta se la distribuzione di probabilità è concentrata solo su k punti $x_i, i = 1, \dots, k$ con la condizione:

$$\sum_{i=1}^k P(x = x_i) = 1 \tag{7.2}$$

In caso opposto la vc si dice continua. Analogamente alla vc discreta ad una dimensione si potrà rappresentare una vc discreta ad n dimensioni con una tabella n -dimensionale.

Nel caso di vc doppia ad esempio si può costruire la tabella:

\underline{x}	x_1^1	x_1^2	\dots	x_1^k	$P_{ij} = P(x_1 = x_1^i, x_2 = x_2^i)$
x_2^1	p_1^1	p_1^2	\dots	p_1^k	
x_2^2	p_2^1	p_2^2	\dots	p_2^k	
x_2^b	p_b^1	p_b^2	\dots	p_b^k	

La vc discreta è sempre assimilabile alla variabile statistica, sostituendo alle p_{ij} le frequenze relative f_{ij} :

$$f_{ij} = \frac{N_{ij}}{N}$$

Una distribuzione di probabilità viene chiamata variabile casuale quando è definita la probabilità per ogni insieme del tipo:

$$\{x_1 \leq x_{01}; \dots x_n < x_{0n}\}$$

$$P(x_1 \leq x_{01}; \dots x_n < x_{0n}) = F(x_{01}; x_{02}; \dots x_{0n}) = F(\underline{x}_0)$$

Anche in questo caso possiamo definire la funzione densità di probabilità della variabile casuale \underline{x} se esiste, attraverso il limite:

$$f(\underline{x}) = \lim_{\rho \rightarrow 0} \frac{P(A)}{\omega(A)} \tag{7.3}$$

dove $\omega(A)$ è la misura dell'insieme A e ρ è il suo «diametro» che tende a zero in \mathbf{R}^n attorno al punto \underline{x} .

La 7.3 può essere riscritta con:

$$f(\underline{x}) = \frac{dP(\underline{x})}{dV(\underline{x})} \tag{7.4}$$

dove $dV(\underline{x})$ è un elemento di volume in \mathbf{R}^n attorno a \underline{x} . Dalla definizione precedente si ha:

$$P(x \in A) = \int_A f(\underline{x}) dV(\underline{x}) \tag{7.5}$$

e la funzione di distribuzione

$$F(x_{01}, x_{02}, \dots, x_{0n}) = \int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{\infty} f(\underline{x}) dx_n \tag{7.6}$$

derivando la 7.6 si ricava:

$$f(\underline{x}) = \frac{\partial^n F(\underline{x})}{\partial x_1 \dots \partial x_n} \tag{7.7}$$

Esempio 1

In un'urna sono contenute due palline bianche (b, B) e due nere (n, N). La variabile casuale discreta che descrive l'estrazione in blocco delle due palline e la relativa probabilità sono¹:

¹ Si ricorda che gli esempi sono tratti dal già citato testo di F. Sansò.

	b	B	n	N
b	/	bB	bn	bN
Bn	Bb	/	Bn	BN
N	nb	nB	/	nN
/	Nb	NB	NB	/

1 ^A ESTRAZIONE			
0	1/12	1/12	1/12
1/12	0	1/12	1/12
1/12	1/12	0	1/12
1/12	1/12	1/12	0

→

2^A
estrazione

Nell'ipotesi di due estrazioni successive con sostituzione (reintegrazione) invece la vc sarà:

	b	B	n	N
b	bb	bB	bn	bN
Bn	Bb	BB	Bn	BN
N	nb	nB	nn	nN
/	Nb	NB	NB	NN

1 ^A ESTRAZIONE			
1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16

→

2^A
estrazione

Esempio 2

Osservando un gran numero di tiri al bersaglio possiamo dire quanto segue:

- a. in ogni zona del bersaglio i colpi tendono a distribuirsi uniformemente a parità di distanza dal centro
- b. contando i punteggi si è visto che, indicando con r la distanza dal centro

$$P[r \in dr(r_0)] = \frac{r_0}{\sigma^2} e^{-\frac{r_0^2}{2\sigma^2}} dr \tag{7.8}$$

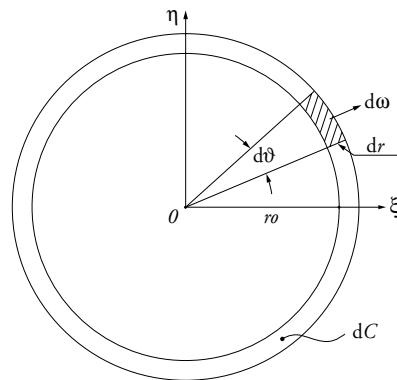


Fig. 7.1 – Distribuzione bidimensionale.

La costante σ^2 è un parametro di bravura del tiratore.

Si vuole trovare la distribuzione bidimensionale dei tiri (figura 7.1).

Notiamo che la 7.8 fornisce la probabilità che $P[\xi, \eta] \in dC$ con dC elemento di corona circolare attorno ad r_0 .

Siccome in dC la probabilità è uniformemente distribuita, allora:

$$P[(\xi, \eta) \in d\omega] = P[x \in dC] \frac{d\omega}{dC} = P[x \in dC] \frac{d\vartheta}{2\pi} = \frac{1}{\sigma^2} e^{-\frac{r_0^2}{2\sigma^2}} r dr \frac{d\vartheta}{2\pi}$$

Per la definizione di densità di probabilità:

$$f(\xi, \eta) = \frac{P[(\xi, \eta) \in d\omega]}{d\omega} = \frac{P[(\xi, \eta) \in d\omega]}{r dr d\vartheta}$$

$$f(\xi, \eta) = \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} = \frac{1}{2\pi\sigma^2} e^{-\frac{\xi^2 + \eta^2}{2\sigma^2}} \quad 7.9$$

La 7.9 rappresenta l'equazione della distribuzione normale a due dimensioni.

7.1 DISTRIBUZIONI MARGINALI

Lo scopo dell'introduzione delle distribuzioni marginali e delle distribuzioni condizionate è, ai nostri fini, capire se e quando due variabili casuali sono fra loro indipendenti.

Consideriamo l'evento A :

$$A = \{x_1 \in dx_1(x_{01}); -\infty < x_2 < \infty; \dots -\infty < x_n < \infty\}$$

È facile intuire che la classe di questi eventi dipende solo dalla variabile casuale x_1 e, nel cercare la probabilità dell'evento A_i , domandiamo qual è la probabilità che x_1 stia in dx_1 qualunque valore assunto per $x_2 \dots x_n$. Da una distribuzione n -dimensionale si genera cioè una distribuzione mono-dimensionale ed una corrispondente vc x_1 tale che:

$$P[x_1 \in dx_1] = P[x \in A] = dx_1 \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} dx_3 \dots \int_{-\infty}^{\infty} dx_n f(x_{01}, x_2, \dots, x_n)$$

Questa vc è detta marginale della \underline{x} ed ha densità di probabilità:

$$f_{x_1}(x_{01}) = \frac{P[x \in A]}{dx_1} = \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} dx_3 \dots \int_{-\infty}^{\infty} dx_n f(x_{01}, x_2, \dots, x_n)$$

ricordando la definizione di densità di probabilità 6.23 come derivata dalla funzione F si ha:

$$f_{x_1}(x_{01}) = \frac{\partial}{\partial x_1} F(x_{01}, +\infty, +\infty, \dots, +\infty) \quad 7.10$$

Una vc n -dimensionale avrà n marginali mono-dimensionali.

Oltre alle distribuzioni marginali ad una componente si possono anche introdurre distribuzioni marginali di insiemi di componenti: (x_1, x_2) , (x_1, x_3) ecc. Ad esempio:

$$f_{x_1, x_2}(x_1, x_2) = \int_{-\infty}^{\infty} dx_3 \dots \int_{-\infty}^{\infty} dx_n f(x_1, x_2, \dots, x_n)$$

che, integrata, fornisce la probabilità che un certo gruppo di componenti (x_1, x_2) appartengano ad un certo elemento di volume dV_2 per qualunque valore assunto dalle altre componenti.

7.2 DISTRIBUZIONI CONDIZIONATE

Ci si chiede qual è la probabilità che m variabili, ad esempio $(x_1 \dots x_m)$ stiano in un elemento di volume dV_m , mentre le altre $(x_{m+1} \dots x_n)$ sono certamente vincolate ad un elemento di volume dV_{n-m} .

I due eventi A e B sono:

$$A\{(x_1 \dots x_m) \in dV_m\}; \quad B\{(x_{m+1} \dots x_n) \in dV_{n-m}\}$$

Si desidera calcolare $P[A|B]$ che vale secondo la 6.7:

$$P[A|B] = \frac{P[AB]}{P[B]} = \frac{\int_{dV_m} \int_{dV_{n-m}} f_x(\underline{x}) dV_m dV_{n-m}}{\int_{dV_{n-m}} \int_{R_m} f_x(x_1 \dots x_m, x_{m+1} \dots x_n) dV_m}$$

$$P[A|B] = \frac{\int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{\infty} dx_m f_x(x_1 \dots x_m, x_{m+1} \dots x_n)}{\int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{\infty} dx_m f_x(x_1 \dots x_m, x_{m+1} \dots x_n)}$$

Tale distribuzione di probabilità genera una densità di probabilità per le variabili $(x_1 \dots x_m)$ per qualunque valore delle rimanenti variabili $(x_{m+1} \dots x_n)$ che vale:

$$f_{x_1 \dots x_m | x_{m+1} \dots x_n}(x_1 \dots x_m | x_{m+1} \dots x_n) = \frac{f_x(x_1 \dots x_m, x_{m+1} \dots x_n)}{\int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{\infty} dx_m f_x(x_1 \dots x_m, x_{m+1} \dots x_n)}$$

$$f_{x_1 \dots x_m | x_{m+1} \dots x_n}(x_1 \dots x_m | x_{m+1} \dots x_n) = \frac{f_x(x_1 \dots x_m, x_{m+1} \dots x_n)}{f_{x_{m+1} \dots x_n}(x_{m+1} \dots x_n)} \quad 7.11$$

7.3 INDIPENDENZA STOCASTICA

Leggi relative alle distribuzioni

Ricordando le 6.8 due eventi si definiscono stocasticamente indipendenti se:

$$P[A|B] = P[A] \tag{6.8}$$

Se ci limitiamo ad esaminare un elemento di volume dV_m :

$$P[A] = P[(x_1 \dots x_m) \in dV_m] = f_{x_1 \dots x_m}(x_1 \dots x_m) dV_m$$

si ha allora che, nel caso di eventi indipendenti, la 7.11 deve essere uguale anche a $f_{x_1 \dots x_m}(x_1 \dots x_m)$, cioè a dire:

$$f(x_1 \dots x_m | x_{m+1} \dots x_n) = f_x(x) = f_{x_1 \dots x_m}(x_1 \dots x_m) f_{x_{m+1} \dots x_n}(x_{m+1} \dots x_n) \tag{7.12}$$

Se ciò è verificato le variabili casuali $(x_1 \dots x_m)$ sono stocasticamente indipendenti dalle rimanenti $(x_{m+1} \dots x_n)$.

Se, al contrario, la densità di probabilità totale $f_x(\underline{x})$ può essere fattorizzata nel prodotto:

$$f_x(\underline{x}) = \phi(x_1 \dots x_m) \psi(x_{m+1} \dots x_n) \tag{7.13}$$

le prime variabili sono indipendenti dalle seconde.

Si nota che i termini al secondo membro sono proporzionali alle marginali. Si arriva così al teorema:

Condizione necessaria e sufficiente affinché $(x_1 \dots x_m)$ siano stocasticamente indipendenti da $(x_{m+1} \dots x_n)$ e viceversa, è che la densità di probabilità congiunta si spacchi nel prodotto delle due marginali:

$$f_x(\underline{x}) = f_{x_1 \dots x_m}(x_1 \dots x_m) f_{x_{m+1} \dots x_n}(x_{m+1} \dots x_n) \tag{7.14}$$

Ne segue un facile corollario:

Condizione necessaria e sufficiente affinché le n componenti di una vc n -dimensionale siano tutte tra loro indipendenti è che la densità di probabilità congiunta si spacchi nel prodotto delle n -marginali:

$$f_x(\underline{x}) = f_{x_1}(x_1) f_{x_2}(x_2) \dots f_{x_n}(x_n) \tag{7.15}$$

Si noti, a proposito, che la 7.9 che rappresenta la variabile di Gauss a due dimensioni può rappresentarsi anch'essa dal prodotto:

$$f(\xi \cdot \eta) = f(\xi) \cdot f(\eta) = \frac{1}{\sqrt{2\pi|\sigma|}} e^{-\frac{1}{2}\left(\frac{\xi}{|\sigma|}\right)^2} \cdot \frac{1}{\sqrt{2\pi|\sigma|}} e^{-\frac{1}{2}\left(\frac{\eta}{|\sigma|}\right)^2}$$

7.4 VARIABILI CASUALI FUNZIONI DI ALTRE VARIABILI CASUALI

Trasformazione di variabili

Supponiamo che sia data una funzione g che trasformi variabili da \mathbb{R}^n a \mathbb{R}^m :

$$\underline{y} = g(\underline{x}) \tag{7.16}$$

(g è un vettore di funzioni).

Si può dimostrare che, a partire da una distribuzione di probabilità in \mathbb{R}^n possiamo costruirne una in \mathbb{R}^m così fatta:

Sia $dV_m(Y_0)$ un'elemento di volume di \mathbb{R}^m in un intorno di Y_0 , e sia $A(Y_0)$ l'immagine inversa di $dV_m(Y_0)$, vale a dire l'insieme di:

$$\forall x \in \mathbb{R}^n / g(x) \in dV_m(Y_0)$$

Si pone:

$$P[Y \in dV_m(Y_0)] = P(\underline{x} \in A(Y_0)) \tag{7.17}$$

ammesso che il secondo termine sia misurabile.

Dunque da una variabile casuale \underline{x} (a destra dell'uguale) possiamo costruirne una seconda (a sinistra dell'uguale).

Ci si chiede: conoscendo la distribuzione di \underline{x} come sarà distribuita la variabile \underline{y} ?

I casi da prendere in considerazione sono tre:

$$m < n;$$

$$m = n;$$

$$m > n.$$

Escludiamo subito il caso $m > n$, infatti, se $g(x)$ è differenziabile l'insieme dei valori argomentali $Y = g(x) / x \in \mathbb{R}^n$ è un insieme in \mathbb{R}^m , ma avrebbe misura nulla: non ci interessa per il trattamento delle misure analizzare distribuzioni singolari.

Nel caso in cui $n=m$, se lo jacobiano J della funzione non è nullo, si ha una cosiddetta trasformazione regolare:

$$J(\underline{g}) = \left| \frac{\partial \underline{g}}{\partial \underline{x}} \right| = \det. \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial g_n}{\partial x_1} & \dots & \frac{\partial g_n}{\partial x_n} \end{vmatrix} \neq 0 \quad \forall x \in \mathbb{R}^n$$

ciò ci permette di dire che esiste anche la relazione inversa che porta da \underline{y} a \underline{x} .

Sia allora $dV_n(y)$ un elemento di volume attorno ad \underline{y} e $dV_n(x)$ l'elemento di volume corrispondente attorno ad \underline{x} .

Il primo intorno lo otteniamo applicando ad \underline{x} la trasformazione g , cioè è l'intorno:

$$dV_n(\underline{y}) = g(dV_n(\underline{x})) \quad 7.18$$

Per la definizione della probabilità ad n -dimensioni si ha poi l'equazione:

$$P[Y \in dV_n(\underline{y})] = P[X \in dV_n(\underline{x})] \quad 7.19$$

e, per la definizione di densità di probabilità:

$$f_y(\underline{y}) dV_n(\underline{y}) = f_x(\underline{x}) dV_n(\underline{x})$$

cioè:

$$f_y(\underline{y}) = \frac{f_x(\underline{x})}{\left[\frac{dV_n(\underline{y})}{dV_n(\underline{x})} \right]} \quad 7.20$$

Ma la derivata al denominatore è qualcosa di già noto, infatti è lo Jacobiano di \underline{g} , $J(\underline{g})$:

$$\left| \det. \left[\frac{\partial \underline{g}}{\partial \underline{x}} \right] \right| = \left| \frac{\partial \underline{g}}{\partial \underline{x}} \right| = \frac{dV_n(\underline{y})}{dV_n(\underline{x})} \quad 7.21$$

e allora la 7.20 si trasforma in:

$$f_y(\underline{y}) = \frac{f_x(\underline{x})}{\left| \frac{\partial \underline{g}}{\partial \underline{x}} \right|} \quad 7.22a$$

dove:

$$\underline{x} = \underline{g}^{-1}(\underline{y}) \quad 7.22b$$

Esempio di applicazione della trasformazione ad un caso lineare

Sia data una trasformazione lineare e regolare da \mathbf{R}^n a \mathbf{R}^m ¹

$$\underline{y} = \underline{A} \underline{x} + \underline{b} \quad 7.23$$

con:

$$|A| = \det.A = \left| \frac{\partial \underline{g}}{\partial \underline{x}} \right| \neq 0$$

¹ Qui di seguito indicheremo di tanto in tanto con doppia sottolineatura le matrici e con singola i vettori. Questa notazione è usata per rendere più chiaro il discorso all'inizio di un problema ed è tralasciata se il senso della formula è univoco, od in genere, per brevità, all'interno di una dimostrazione già avviata.

Si ha:

$$\underline{x} = \underline{A}^{-1}(\underline{y} - \underline{b}) \quad 7.24$$

$$f_y(\underline{y}) = \frac{f_x(\underline{A}^{-1}(\underline{y} - \underline{b}))}{|A|} \quad 7.25$$

Sia la funzione di distribuzione $f_x(\underline{x})$, ad esempio, il prodotto di n normali standardizzate, tali che:

$$f_x(\underline{x}) = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \dots \frac{1}{\sqrt{2\pi}} e^{-x_n^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\sum x_i^2/2} \quad 7.26a$$

che può essere anche scritta come:

$$f_x(\underline{x}) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{(x^T x)}{2}} \quad 7.26b$$

Dalla 7.24 ricaviamo:

$$x^T = [A^{-1}(\underline{y} - \underline{b})]^T = (\underline{y} - \underline{b})^T (A^{-1})^T$$

Si ottiene infine dalla 7.25:

$$f_y(\underline{y}) = \frac{1}{(2\pi)^{n/2} |A|} e^{-\frac{1}{2}(\underline{y} - \underline{b})^T (A^{-1})^T A^{-1} (\underline{y} - \underline{b})} \quad 7.27$$

Esaminiamo l'esponente della 7.27.

Definita A^2 una matrice reale, simmetrica e positiva, si dimostra che è sempre possibile scomporla nel prodotto:

$$A^2 = A^T A = A A^T \quad 7.28$$

di modo che la 7.27 diviene:

$$f_y(\underline{y}) = \frac{1}{(2\pi)^{n/2} |A|} e^{-\frac{1}{2}(\underline{y} - \underline{b})^T (A^2)^{-1} (\underline{y} - \underline{b})} \quad 7.29$$

La 7.29 rappresenta la forma nella quale è possibile scrivere la funzione densità di probabilità di una qualsiasi variabile normale n -dimensionale non standardizzata e rappresenta anche, con la 7.23 e la 7.28 la via da seguire per la standardizzazione. Questi concetti saranno ripresi ed estesi in seguito.

Esaminiamo infine il caso di una trasformazione da \mathbf{R}^n a \mathbf{R}^m con $m < n$, cioè:

$$\begin{cases} y_1 = g_1(x_1 \dots x_n) \\ y_m = g_m(x_1 \dots x_n) \end{cases} \quad 7.30$$

Ad un elemento di volume $dV_m(\underline{y})$ può corrispondere un insieme di $\underline{x} = A_x(dV_m)$ che non ha misura finita:

$$dV_m(\underline{y}) = g[A_x(dV_m)]$$

Ponendo:

$$P[Y \in dV_m(Y_0)] = P[x \in A_{X_0}]$$

si ha:

$$f_Y(\underline{y}) = \frac{1}{dV_m(\underline{y})} \int_{A_x(dV_m)} f_x(\underline{x}) dV_n(\underline{x}) \quad 7.31$$

Oltre alla 7.31, se non intervengono ulteriori ipotesi, non si può in questo caso dire altro.

7.5 MOMENTI DI VARIABILI n -DIMENSIONALI

Anche per le variabili casuali n -dimensionali possono generalizzarsi i concetti visti ad una dimensione.

Se esiste la media della variabile casuale n -dimensionale \underline{x} questa è per definizione un vettore n -dimensionale $\underline{\mu}_x$ dato da:

$$\underline{\mu}_x = M[x] = \int_{R^n} dV_n(x) f_x(\underline{x}) \bullet \underline{x} \quad 7.32$$

dove il simbolo \bullet sta per prodotto scalare. La componente i -esima di $\underline{\mu}_x$ vale:

$$\mu_{x_i} = M[x_i] = \int_{R^n} dV_n(x) x_i f_x(\underline{x}) \quad 7.33$$

dalla 7.32 si nota che per calcolare μ_{x_i} basta conoscere la distribuzione marginale di x_i , infatti:

$$\begin{aligned} \mu_{x_i} &= \int_{\beta^n} dx_i dV_{n-1} x_i f_x(\underline{x}) = \int_{-\infty}^{\infty} dx_i x_i \int_{-\infty}^{\infty} dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n f_x(\underline{x}) \\ \mu_{x_i} &= \int_{-\infty}^{\infty} dx_i x_i f_{x_i}(x_i) \end{aligned} \quad 7.34$$

cioè la componente i -esima della media di \underline{x} è uguale alla media della componente i -esima.

Nel caso ad esempio di una variabile statistica doppia $[\underline{x}, \underline{y}]$, rappresentata al solito dalla tabella:

$$\begin{cases} x = x_1, x_2, \dots, x_r \\ y = y_1, y_2, \dots, y_s \end{cases}$$

possiamo, sfruttando la solita analogia, ricavare:

$$\begin{aligned} M[x] &= \bar{x} = \frac{1}{r} \sum x_i \\ M[y] &= \bar{y} = \frac{1}{s} \sum y_j \end{aligned}$$

Teorema della media per variabili casuali n -dimensionali

Sia una trasformazione da \mathbb{R}^n a \mathbb{R}^m , con \underline{x} variabile casuale $\underline{x} \in \mathbb{R}^n$ e \underline{y} variabile $\underline{y} \in \mathbb{R}^m$ per definizione di media, se esiste, si ha:

$$M[\underline{y}] = M_y[g(\underline{x})] = \int_{\mathbb{R}^n} g(\underline{x}) f_x(\underline{x}) d\underline{x} \tag{7.35}$$

In questo caso il teorema della media afferma che:

$$M_y[\underline{y}] = M_x[g(\underline{x})] \tag{7.36}$$

Corollario 1

Nel caso in cui la funzione vettoriale g sia lineare, nel caso cioè in cui:

$$\begin{aligned} \underline{y} &= \underline{A} \underline{x} + \underline{b} \\ \underline{\mu}_y &= \underline{A} \underline{\mu}_x + \underline{b} \end{aligned} \tag{7.37}$$

Corollario 2

se la variabile \underline{x} è ben concentrata in una zona di \mathbb{R}^n attorno alla media $\underline{\mu}_x$ e, nella stessa zona la funzione che lega le due variabili casuali: $y = g(x)$ è lentamente variabile allora:

$$\underline{\mu}_y = g(\underline{\mu}_x) \tag{7.38}$$

in analogia a quanto visto per vc ad una dimensione.

Momenti di ordine (n_1, n_2, \dots, n_k) di una variabile casuale n -dimensionale

Si definiscono momenti di ordine (n_1, n_2, \dots, n_k) di una variabile casuale n -dimensionale gli scalari:

$$\mu_{i_1, i_2, \dots, i_k}^{n_1, n_2, \dots, n_k} = M[x_{i_1}^{n_1}, x_{i_2}^{n_2}, \dots, x_{i_k}^{n_k}] \tag{7.39}$$

Si definiscono momenti centrali i corrispondenti momenti della variabile scarto:

$$\underline{v} = \underline{x} - \underline{\mu}_x$$

Molto spesso tuttavia i momenti più usati sono quelli del secondo ordine che, per definizione indichiamo con:

$$c_{ik} = M[(x_i - \mu_{xi})(x_k - \mu_{xk})] = M[v_i v_k] \quad 7.40$$

Notiamo che per $i=k$ si ha:

$$c_{ii} = \sigma_i^2 = M[(x_i - \mu_{xi})^2] \quad 7.41$$

cioè i momenti centrali del secondo ordine per $i=k$ sono le varianze della componente i -esima di \underline{x} .

I coefficienti c_{ik} per $i \neq k$ si indicano anche con σ_{ik} e sono detti coefficienti di covarianza delle componenti x_i e x_k .

Come evidente dalla 7.40 $c_{ik} = c_{ki}$, la 7.40 e la 7.41 espresse in forma matriciale divengono:

$$C_{xx} = [c_{ik}] = M[(x_i - \mu_{xi})(x_k - \mu_{xk})] = C_{xx} = M[(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^T] \quad 7.42$$

La C_{xx} è detta per ovvi motivi matrice di varianza covarianza o matrice di dispersione ed è simmetrica.

Si può dimostrare, analogamente al caso mono-dimensionale, che:

$$C_{xx} = M[xx^T] - \underline{\mu}_x \underline{\mu}_x^T \quad 7.43$$

Cerchiamo ora un'altra espressione della 7.42 nel caso particolare in cui le componenti di \underline{x} siano fra loro indipendenti.

In questo caso $f(\underline{x})$ può essere scritta come prodotto delle marginali 7.15:

$$f(\underline{x}) = f_{x_1}(x_1) \dots f_{x_n}(x_n)$$

e, osservando che ogni marginale è normalizzata per suo conto, cioè che:

$$\int_{-\infty}^{\infty} f_{x_j}(x_j) dx_j = 1$$

si trova, per $i \neq k$, ricordando la 7.40,

$$\begin{aligned} M[x_i x_k] &= \int x_i x_k f_x(x) dx_1 \dots dx_n \\ &= \left(\int_{-\infty}^{\infty} f_{x_1}(x_1) dx_1 \right) \left(\int_{-\infty}^{\infty} f_{x_2}(x_2) dx_2 \right) \left(\int_{-\infty}^{\infty} x_i f_{x_i}(x_i) dx_i \right) \left(\int_{-\infty}^{\infty} x_k f_{x_k}(x_k) dx_k \right) \\ M[x_i x_k] &= \mu_{x_i} \mu_{x_k} \quad 7.44 \end{aligned}$$

ma, ricordando la 7.43:

$$c_{ik} = M[x_i x_k] - \mu_{x_i} \mu_{k_k}$$

ne deriva che:

$$c_{ik} = \sigma_{ik} = 0 \quad \forall i \neq k \quad 7.45$$

cioè, per componenti di \underline{x} indipendenti, la matrice C_{xx} è diagonale e assume la forma:

$$C_{xx} = \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_n^2 \end{pmatrix} \quad 7.46$$

Si può verificare in molti casi che non è vero viceversa, cioè la forma diagonale di C_{xx} non significa necessariamente che le n -componenti siano fra loro indipendenti.

La propagazione della varianza nel caso lineare ad n -dimensioni

Come nel caso mono-dimensionale ci domandiamo cosa vale la matrice di varianza covarianza di una variabile casuale $y \in \mathbf{R}^m$ funzione di una seconda variabile $x \in \mathbf{R}^n$.

L'ipotesi è che la relazione g sia lineare, cioè $\underline{y} = \underline{A} \underline{x} + \underline{b}$ e che $m \leq n$.

Per il teorema della media:

$$\underline{\mu}_y = \underline{A} \underline{\mu}_x + \underline{b}$$

dunque:

$$(\underline{y} - \underline{\mu}_y) = \underline{A}(\underline{x} - \underline{\mu}_x) \quad 7.47$$

ma per definizione di C_{yy} :

$$C_{yy} = M[(\underline{y} - \underline{\mu}_y)(\underline{y} - \underline{\mu}_y)^T] = M[A(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^T A^T]$$

sfruttando la linearità dell'operatore media, $M[\bullet]$, si ha:

$$C_{yy} = A M[(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^T] A^T = A C_{xx} A^T \quad 7.48$$

È questa la legge di propagazione della varianza nel caso lineare.

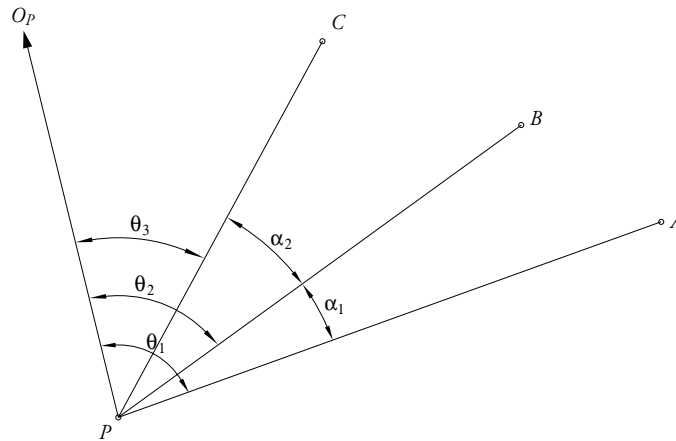
Esercizio 1

Con un teodolite si misurano le direzioni $\vartheta_1, \vartheta_2, \vartheta_3$ che ipotizziamo estratte da una vc a tre dimensioni con media $(\bar{\vartheta}_1, \bar{\vartheta}_2, \bar{\vartheta}_3)$, indipendenti fra di loro e con varianze:

$$\sigma_{\vartheta_1} = \sigma_{\vartheta_2} = \sigma_{\vartheta_3} = \pm 10 \cdot 10^{-4} \text{ gon} = \sigma$$

Si determini, valor medio, varianza e covarianza degli angoli azimutali α_1 e α_2 così definiti:

$$\begin{cases} \alpha_1 = \vartheta_1 - \vartheta_2 \\ \alpha_2 = \vartheta_2 - \vartheta_3 \end{cases}$$

**Fig. 7.2**

L'esercizio è lasciato allo svolgimento del lettore con questo suggerimento: data la matrice

$$C_{\vartheta\vartheta} = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$$

si applichi il teorema della media e la propagazione della varianza da \mathbb{R}^3 a \mathbb{R}^2 .

Esercizio 2

Si calcoli la covarianza fra x e y e le rispettive varianze per la seguente variabile statistica doppia:

$y = \rightarrow$	4	5	9	p_i
$x = \downarrow$				\downarrow
1	0.1	0.2	0.1	0.4
2	0.1	0.2	0	0.3
3	0	0.1	0.1	0.2
4	0	0	0.1	0.1
$q_j \rightarrow$	0.2	0.5	0.3	1

Si ricavano dapprima le frequenze p_i e q_j delle marginali; i valori medi sono ricavati attraverso le frequenze marginali:

$$M_x = \sum_1^{n=4} x_i p_i = 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.2 + 4 \cdot 0.1 = 2$$

$$M_y = \sum_1^{m=3} y_j q_j = 4 \cdot 0.2 + 5 \cdot 0.5 + 9 \cdot 0.3 = 6$$

Per definizione:

$$\sigma_{xy} = \sum_{i=1}^n \sum_{j=1}^m (x_i - M_x)(y_j - M_y) f_{ij}$$

$$\sigma_{xy} = \sum_{i=1}^n \sum_{j=1}^m x_i y_j f_{ij} - \sum_{i=1}^n \sum_{j=1}^m x_i M_y f_{ij} - \sum_{j=1}^m \sum_{i=1}^n y_j M_x f_{ij} + \sum_{i=1}^n \sum_{j=1}^m M_x M_y f_{ij}$$

Al secondo membro il secondo termine vale $-M_y M_x$ ed il terzo vale $-M_x M_y$, infine il quarto vale $M_x M_y$ essendo:

$$\sum_{i=1}^n \sum_{j=1}^m f_{ij} = 1$$

Si ha infine:

$$\sigma_{xy} = \sum_{i=1}^n \sum_{j=1}^m x_i y_j f_{ij} - M_x M_y \tag{7.49}$$

che rappresenta l'estensione della 7.43. Sostituendo infatti x ad y o viceversa si trova:

$$\sigma_x^2 = \sum_{i=1}^n \sum_{i=1}^n x_i^2 - M_x^2$$

Applicando tutto ciò ai dati dell'esercizio si ricava:

$$\sigma_x^2 = \sum_{i=1}^4 x_i^2 p_i - M_x^2 = (1 \cdot 0.4 + 4 \cdot 0.3 + 9 \cdot 0.2 + 16 \cdot 0.1) - 4 = 1$$

$$\sigma_y^2 = \sum_{j=1}^3 y_j^2 q_j - M_y^2 = (16 \cdot 0.2 + 25 \cdot 0.5 + 81 \cdot 0.3) - 36 = 4$$

$$\begin{aligned} \sigma_{xy} &= \sum_{i=1}^4 x_i \sum_{j=1}^3 y_j f_{ij} - M_x M_y = \\ &= 1 \cdot (4 \cdot 0.1 + 5 \cdot 0.2 + 9 \cdot 0.1) + 2 \cdot (4 \cdot 0.1 + 5 \cdot 0.2) + \\ &\quad + 3 \cdot (5 \cdot 0.1 + 9 \cdot 0.1) + 4 \cdot (9 \cdot 0.1) - 12 \end{aligned}$$

$$\sigma_{xy} = 2.3 + 2.8 + 4.2 + 3.6 - 12 = 0.9$$

Si ha allora che:

$$C_{xy} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 4 \end{pmatrix}$$

7.6 LA LEGGE DI PROPAGAZIONE DELLA VARIANZA NEL CASO DI FUNZIONI NON LINEARI

Poniamoci ancora nel caso (n, m) dimensionale in cui $m \leq n$ e sia:

$$\underline{y} = g(\underline{x}) \tag{7.50}$$

una funzione non più lineare della variabile casuale \underline{x} .

Nell'ipotesi che \underline{x} sia ben concentrato attorno alla sua media $\underline{\mu}_x$ ed \underline{y} sia poco variabile attorno a $g(\underline{\mu}_x)$ si può operare la linearizzazione:

$$\underline{y} \cong g(\underline{\mu}_x) + \left[\frac{\partial g}{\partial x} \right] (\underline{x} - \underline{\mu}_x) \tag{7.51}$$

È ora possibile utilizzare le 7.47 e 7.48 ricavate per il caso lineare con le seguenti sostituzioni:

$$\underline{b} = g(\underline{\mu}_x) \tag{7.52}$$

$$\underline{\underline{A}} = \left[\frac{\partial \underline{g}}{\partial \underline{x}} \right] \quad 7.53$$

La matrice A è detta *matrice disegno*. La 7.48 diviene allora:

$$C_{yy} = \left[\frac{\partial \underline{g}}{\partial \underline{x}} \right] C_{xx} \left[\frac{\partial \underline{g}}{\partial \underline{x}} \right]^T \quad 7.54$$

Le matrici C_{xx} e C_{yy} sono sempre strettamente definite positive, cioè (definizione):

$$C_{xx} > 0: \forall \underline{a} \in \mathbf{R}^n / \underline{a}^T C_{xx} \underline{a} > 0 \quad 7.55$$

Si fissi infatti \underline{a} e si consideri $y = \underline{a}^T x$, con y variabile casuale mono-dimensionale; si avrà come logico $\sigma_y^2 \geq 0$ e $\sigma_y^2 > 0$ se x non ha distribuzioni singolari come nell'ipotesi di trasformazioni regolari.

Se C_{xx} è regolare (invertibile) e simmetrica, è sempre poi possibile questa scomposizione:

$$C_{xx} = K^2 = U \Lambda U^T \quad 7.56$$

Con Λ matrice diagonale degli autovalori di C_{xx} ed U matrice ortogonale $U^T U = U U^T = I$ che contiene gli autovettori di C_{xx} . È facile dopo questa ipotesi dimostrare che:

$$K = U \Lambda^{1/2} U^T \quad 7.57$$

La radice quadrata di una matrice diagonale Λ è la matrice i cui elementi valgono $\sqrt{\lambda_i}$.

Esercizio 3

Di un punto P si sono misurate la distanza dall'origine r e l'anomalia ϑ , rappresentate dalle variabili casuali ρ e ϑ con media e sqm seguenti:

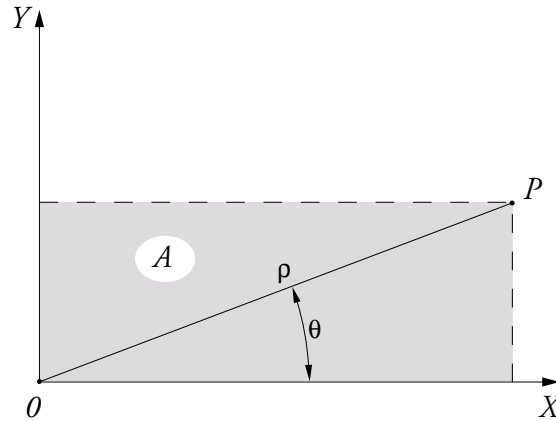
$$\begin{aligned} \bar{\rho} &= 1 \text{ km} & \sigma_{\rho} &= \pm 1 \text{ mm} & (\bar{\rho} &= 10^6 \text{ mm}) \\ \bar{\vartheta} &= \pi/6 & \sigma_{\vartheta} &= \pm 2 \cdot 10^{-6} \text{ (rad)} \end{aligned}$$

Calcolare media e covarianza delle coordinate (x, y) del punto P e media e varianza dell'area A del rettangolo che ha \overline{OP} per diagonale.

La trasformazione g permette di ricavare (x, y) in funzione delle misure dirette (ρ, ϑ) .

(x, y) sono misurabili cioè indirettamente.

$$\eta = \begin{pmatrix} x \\ y \end{pmatrix}; \quad \xi = \begin{pmatrix} \rho \\ \vartheta \end{pmatrix}; \quad C_{\xi\xi} = \begin{pmatrix} \sigma_{\rho}^2 & 0 \\ 0 & \sigma_{\vartheta}^2 \end{pmatrix} = \begin{pmatrix} 1 \text{ mm}^2 & 0 \\ 0 & 4 \cdot 10^{-12} \end{pmatrix}$$


Fig. 7.3

Applicando il teorema della media si ricavano i valori medi:

$$\bar{\eta} = g(\bar{\xi}) = \begin{pmatrix} \bar{\rho} \cos \bar{\vartheta} \\ \bar{\rho} \sin \bar{\vartheta} \end{pmatrix} \quad \begin{aligned} \Rightarrow \mu_x &= 866.025 \text{ mm} \\ \Rightarrow \mu_y &= 500.000 \text{ mm} \end{aligned}$$

Si ricava ora la matrice disegno, calcolandola nell'intorno dei valori medi:

$$\left[\frac{\partial g}{\partial \xi} \right] = \begin{pmatrix} \cos \vartheta & -\rho \sin \vartheta \\ \sin \vartheta & \rho \cos \vartheta \end{pmatrix}; \quad \left[\frac{\partial g}{\partial \xi} \right]_{\mu \xi} = \begin{bmatrix} \sqrt{3}/2 & -10^{-6}/2 \text{ m} \\ 1/2 & \frac{10^6 \sqrt{3}}{2} \text{ m} \end{bmatrix}$$

Si verifica poi se la trasformazione è regolare.

$$\det. \left[\frac{\partial g}{\partial \xi} \right]_{\mu \xi} = \bar{\rho} (\cos^2 \bar{\vartheta} + \sin^2 \bar{\vartheta}) = \bar{\rho} > 0$$

Si applica infine il teorema di propagazione della varianza:

$$C_{\eta\eta} = \begin{pmatrix} \sqrt{3}/2 & -10^6/2 \\ 1/2 & \frac{10^6 \sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 4 \cdot 10^{-12} \end{pmatrix} \begin{pmatrix} \sqrt{3}/2 & 1/2 \\ -10^6/2 & \frac{10^6 \sqrt{3}}{2} \end{pmatrix} = \begin{pmatrix} 1.75 & -1.30 \\ -1.30 & 3.25 \end{pmatrix}$$

Per rispondere alle ultime due domande applichiamo ancora il teorema della media alla misura indiretta – superficie A – funzione delle due misure dirette ρ e ϑ :

$$\bar{A} = \bar{\rho}^2 \sin \bar{\vartheta} \cos \bar{\vartheta} \Rightarrow \bar{A} = 0.433 \cdot 10^6 \text{ m}^2$$

Ed applicando il principio di propagazione della varianza si ricava:

$$\sigma_A^2 = \left(\bar{\rho} \sin \frac{2\bar{\vartheta}}{2} \right)^2 \sigma_\rho^2 + (\bar{\rho}^2 \cos 2\bar{\vartheta})^2 \sigma_\vartheta^2 = \left(10^{12} \cdot \frac{3}{4} \cdot 1 \right) \frac{1 + 10^{24}}{4} \cdot \frac{1}{4} \cdot 4 \cdot 10^{-12}$$

$$\sigma_A = \pm 1.323 \text{ m}^2$$

Si lascia come esercizio ricavare quest'ultimo risultato a partire dalla relazione $A = xy$, con $C_{\eta\eta}$ ricavata come sopra.

La propagazione della varianza da n dimensioni ad una dimensione

L'esercizio precedente è un caso particolare nel quale è possibile ricavare una formula semplificata rispetto alle 7.48 e 7.54.

Nel caso di trasformazione da n -dimensioni ad una dimensione, l'unica incognita è la varianza σ_y^2 .

Partendo dalla relazione:

$$y = f(x_1, x_2, \dots, x_n) \tag{7.58}$$

con C_{xx} matrice di varianza covarianza di \underline{x} . La 7.54 diviene:

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \dots + 2\left(\frac{\partial f}{\partial x_1} \cdot \frac{\partial f}{\partial x_2}\right) \sigma_{12} + \dots + 2\left(\frac{\partial f}{\partial x_i} \cdot \frac{\partial f}{\partial x_k}\right) \sigma_{ik} \tag{7.59}$$

A conclusione di questa prima parte del trattamento statistico delle misure si propongono questi esercizi.

Esercizio 1

Sia data una v.s. x

$$x = \begin{cases} 10 \div 12 & 12 \div 15 & 15 \div 20 & 20 \div 30 & 30 \div 50 \\ 0.04 & 0.18 & 0.40 & 0.20 & 0.18 \end{cases}$$

- Calcolare:
- l'istogramma
 - la funzione di distribuzione
 - la media, la mediana (è l'ascissa per cui $P = 1/2$)
 - la varianza
 - verificare il teorema di Tchebjcheff tra $(\mu-10)$ e $(\mu+10)$.

Esercizio 2

Sia di una v. casuale $f(x) = k$.

Calcolare:

- k
- \bar{x}
- σ_x^2
- verificare il Teorema di Tchebjcheff

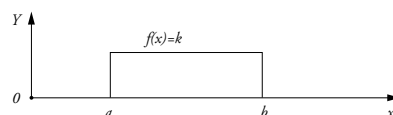


Fig.7.4

Esercizio 3

Sia di una v. casuale $f(x) = kx$.

Calcolare:

- k
- \bar{x}
- σ_x^2
- verificare il Teorema di Tchebjeff

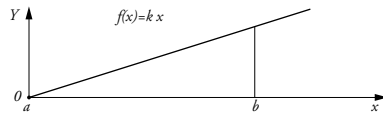


Fig.7.5

Esercizio 4

Trasformazioni di variabili casuali. Sia:

$$\begin{cases} f_x(x) = \text{cost} \\ y = x^2 \end{cases}$$

trovare:

- k
- $f_y(y)$
- $M[y]$ e $\sigma^2(y)$
- verificare il Teorema di Tchebjeff

Fare lo stesso esercizio per: $y = \log(x)$; verificare se $M[y] = g(M[x])$.

Esercizio 5

Di un triangolo si sono misurati direttamente a, b e l'angolo compreso γ .

Dati $\sigma_a, \sigma_b, \sigma_\gamma$, calcolare la superficie media S ed il suo sigma.

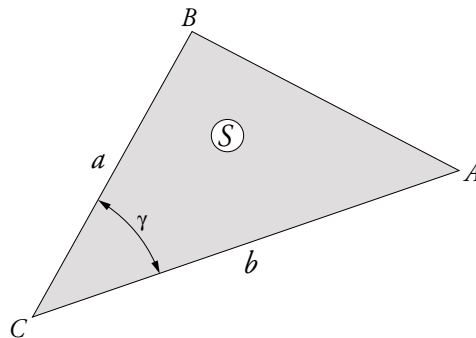


Fig.7.6

7.7 INDICE DI CORRELAZIONE LINEARE

Supponiamo che \bar{x} e \bar{y} siano variabili casuali ad n -dimensioni e che siano fra loro indipendenti. Si avrà allora:

$$\sigma_{xy} = M[xy] - \mu_x \mu_y = 0 \quad 7.60$$

Ipotizziamo ora invece che y sia funzionalmente dipendente da x , e che lo sia inoltre in modo lineare:

$$\underline{y} = \underline{A}\underline{x} + \underline{b}$$

Ne deriva che, come già visto:

$$(y - \mu_y) = A(x - \mu_x)$$

Cerchiamo ora le covarianze tra x ed y :

$$C_{xyxy} = M[(x - \mu_x)^T(y - \mu_y)] = M[(x - \mu_x)^T A(x - \mu_x)]$$

cioè:

$$C_{xyxy} = AC_{xx} \quad 7.61$$

Ora poniamoci nel caso di x ed y ad una componente; nell'ipotesi di indipendenza della 7.59 si avrà che $\sigma_{xy} = 0$, mentre nell'ipotesi che ha portato alla 7.61 $\sigma_{xy} = a^2 \sigma_x^2$; inoltre, siccome $y = ax + b$, applicando la propagazione della varianza si ricava $\sigma_y^2 = a^2 \sigma_x^2$, cioè $\sigma_y = |a| \sigma_x$.

Definiamo indice di correlazione lineare di x ed y lo scalare ρ_{xy} :

$$\rho_{xy} = (\text{def.}) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad 7.62$$

Nella seconda ipotesi di dipendenza lineare si ha:

$$\rho_{xy} = \frac{a \sigma_x^2}{\sigma_x |a| \sigma_y} = \pm 1$$

Nella prima ipotesi 7.60 di indipendenza si può facilmente verificare che:

$$\rho_{xy} = \frac{0}{\sigma_x \sigma_y} = 0$$

Questo parametro varia dunque nell'intervallo ± 1 e vale zero per variabili casuali fra loro indipendenti.

Si osservi che, viceversa, se $\rho_{xy} = 0$ le due variabili casuali si dicono incorrelate ma non è detto che siano indipendenti.

La figura 7.7 mostra un caso di distribuzione di densità di probabilità di variabili dipendenti ma incorrelate.

ρ_{xy} è un parametro molto molto utilizzato grazie a queste sue proprietà:

- è invariante in modulo per trasformazioni lineari, cioè non cambia se cambiano linearmente le unità di misura di x e y .
- se x e y sono variabili indipendenti $\rho_{xy} = 0$; se al contrario sono linearmente dipendenti, assume valore $\rho_{xy} = \pm 1$; $+1$ per $a > 0$, e -1 per $a < 0$, si ha cioè $\sigma_{xy} = \pm \sigma_x \sigma_y$.

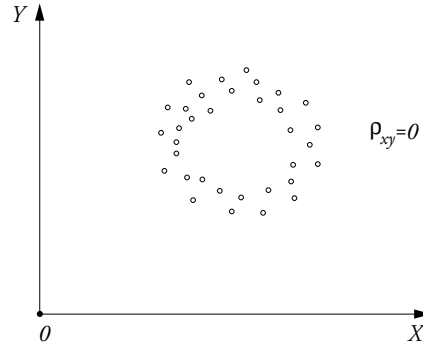


Fig. 7.7 – Variabili incorrelate ma non indipendenti.

Si può dimostrare che per una variabile doppia non ordinata vale:

$$\rho_{xy} = \frac{N \sum x_i y_i - \sum x_i \cdot \sum y_i}{\sqrt{\left[N \sum x_i^2 - \left(\sum x_i \right)^2 \right] \left[N \sum y_i^2 - \left(\sum y_i \right)^2 \right]}} \quad 7.63$$

7.8 PROPRIETÀ DELLE VARIABILI NORMALI AD n -DIMENSIONI

Ricordiamo l'espressione 7.26b della variabile normale n -dimensionale con:

$$\begin{aligned} M[x] &= 0 \\ C_{x_i x_i} &= \text{diag}(\sigma_{x_i}^2) = 1 \\ C_{x_i x_k} &= 0 \end{aligned} \quad 7.64$$

cioè:

$$C_{xx} = I \quad 7.65$$

Supponiamo ancora di eseguire una trasformazione lineare del tipo 7.23: $y = Ax + b$ ma ora ipotizziamo che la matrice A possa essere scritta in questo modo:

$$A = \Lambda^{1/2} U \quad 7.66$$

con U matrice ortogonale e $\Lambda^{1/2}$ matrice diagonale. Ricordando la 7.29 si ha:

$$f_y(y) = \frac{1}{2\pi^{n/2} |\Lambda U|} e^{-\frac{1}{2} (y-b)^T (U \Lambda U^T)^{-1} (y-b)} \quad 7.67$$

Ora, ricordando la 7.56 che esprime la forma di una qualsiasi matrice regolare simmetrica possiamo sfruttare il risultato a ritroso per standardizzare la variabile casuale y .

La trasformazione inversa sarà dunque:

$$x = \Lambda^{-1/2} U^T(y - b) = C_{yy}^{-1/2}(y - b) \quad 7.68$$

Questa operazione si chiama appunto standardizzazione della variabile casuale y , la quale ha media $\mu_y = b$ e matrice di varianza covarianza C_{yy}

Per dichiarare che y appartiene ad una distribuzione normale con tali medie e varianze si scrive:

$$y = N[b, C_{yy}]$$

Vediamo due proprietà delle variabili casuali normali:

1. Il concetto di correlazione ed indipendenza stocastica si equivalgono.
2. Tutte le trasformazioni lineari trasformano variabili normali in variabili normali; cioè se:

$$x = N[\mu_x, C_{xx}]$$

e se:

$$y = Ax + b$$

allora, ammesso che $m \leq n$ e che il rango di A sia pieno, $r(A) = m$:

$$y = N[A\mu_x + b; AC_{xx}A^T].$$

Si osserva che la variabile:

$$(x - \mu)^T C_{xx}^{-1}(x - \mu) = z^T z = \sum_{i=1}^n z_i^2 = \chi_n^2 \quad 7.69$$

è una variabile casuale χ_n^2 a n gradi di libertà; ciò consente di trovare attorno al vettore media $\mu_x \in \mathbb{R}^n$ una regione simmetrica nella quale sia contenuta una prefissata probabilità $P = p$ cioè:

$$P[(x - \mu)^T C_{xx}^{-1}(x - \mu) \leq \chi_n^2] = p$$

I valori più usati sono $p = 50\%$, $p = 90\%$.

La regione:

$$(x - \mu)^T C_{xx}^{-1}(x - \mu) \leq \overline{\chi}_n^2 \quad 7.70$$

risulta essere un iper-ellissoide. Per $n = m = 2$ ad esempio, si noti che:

$$\det(C_{xx}) = \sigma_x^2 \sigma_y^2 - \sigma_{xy}^2 = \sigma_x^2 \sigma_y^2 - \rho^2 \sigma_x^2 \sigma_y^2 = \sigma_x^2 \sigma_y^2 (1 - \rho^2)$$

e dunque:

$$C_{xx}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_x^2} & \frac{-\sigma_{xy}}{\sigma_x^2 \sigma_y^2} \\ \frac{-\sigma_{xy}}{\sigma_x^2 \sigma_y^2} & \frac{1}{\sigma_y^2} \end{pmatrix}$$

essendo:

$$\sigma_{xy} = -\rho \sigma_x \sigma_y$$

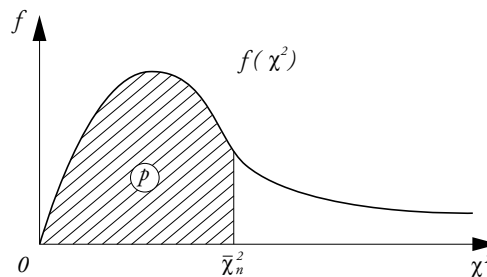


Fig. 7.8 – Uso della variabile χ^2 .

$$P(\chi_n^2 \leq \bar{\chi}_n^2) = p$$

$$\frac{(x - \mu_x)^2}{\sigma_x^2} - \frac{2(x - \mu_x)(y - \mu_y)\sigma_{xy}}{\sigma_x^2 \sigma_y^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} = \chi^2(1 - \rho^2) \quad 7.71$$

Si nota con facilità che la 7.71 è un'ellisse, nel caso in cui $\rho = 0$ e $\sigma_{xy} = 0$ ed ha centro in (μ_x, μ_y) .

Dalla 7.71 si nota pure che per una opportuna rotazione di assi l'ellisse ha equazione del tipo:

$$\frac{(\xi - t)^2}{a^2} + \frac{(\eta - u)^2}{b^2} = \bar{\chi}_2^2$$

in tal caso $\rho_{xy} = 0$. Cerchiamo dunque questa rotazione.

Sia (u, v) una variabile normale doppia con matrice di dispersione $C_{uv} = C$:

$$C = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$$

Vogliamo trovare, se possibile, dopo una rotazione degli assi nel piano (u, v) , una nuova variabile normale doppia le cui componenti siano incorrelate ($\sigma_{xy} = 0$).

La trasformazione sarà in genere la rotazione del tipo:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}; \quad \begin{pmatrix} x \\ y \end{pmatrix} = R \begin{pmatrix} u \\ v \end{pmatrix}$$

Si avrà, applicando la legge della propagazione della varianza:

$$C_{xy} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} = RC_{uv}R^T$$

e, sviluppando i prodotti si ottiene:

$$\sigma_x^2 = \sigma_u^2 \cos^2 \alpha - 2\sigma_{uv} \cos \alpha \sin \alpha + \sigma_v^2 \sin^2 \alpha \quad 7.72$$

$$\sigma_y^2 = \sigma_u^2 \sin^2 \alpha + 2\sigma_{uv} \sin \alpha \cos \alpha + \sigma_v^2 \cos^2 \alpha \quad 7.73$$

$$\sigma_{xy} = (\sigma_u^2 - \sigma_v^2) \cos \alpha \sin \alpha + \sigma_{uv} (\cos^2 \alpha - \sin^2 \alpha)$$

imponendo $\sigma_{xy} = 0$, e utilizzando le formule:

$$\cos \alpha \sin \alpha = \frac{\sin 2\alpha}{2}$$

$$\cos 2\alpha = (\cos^2 \alpha - \sin^2 \alpha)$$

ricaviamo:

$$\operatorname{tg} 2\alpha = \left(\frac{2\sigma_{uv}}{\sigma_v^2 - \sigma_u^2} \right) \quad 7.74$$

Ricavata la rotazione α si sostituisce nelle 7.72 e 7.73 e si ricavano i valori σ_x, σ_y .

Si dimostra che questi valori sono rispettivamente i valori di massimo e di minimo σ^2 , e si indicano perciò rispettivamente con σ_I, σ_{II} .

Tali valori si chiamano semiassi principali dell'ellisse d'errore, o dell'ellissoide od iperellissoide nel caso in cui fossimo nello spazio a più di due dimensioni.

Estendendo il risultato ad n -dimensioni si può infatti ancora dimostrare che è possibile trovare una matrice di rotazione U tale che attraverso il cambiamento di variabile dovuto alla matrice U :

$$y = Ux$$

$$C_{yy} = \operatorname{diag}(c_{ii})$$

Per ulteriori approfondimenti si veda l'appendice A. Per una variabile bidimensionale, nell'ipotesi semplificativa $\mu_x = \mu_y = 0$ la 7.70 diviene:

$$(x \quad y) \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \text{cost}$$

Anche questa curva rappresenta un'ellisse. I semiassi principali sono rappresentati dagli autovalori della matrice C_{xx} (non di C_{xx}^{-1}) che ricaviamo da:

$$|C - \lambda I| = 0$$

cioè:

$$\begin{vmatrix} \sigma_x^2 - \lambda & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 - \lambda \end{vmatrix} = 0$$

$$(\sigma_x^2 - \lambda)(\sigma_y^2 - \lambda) - \sigma_{xy}^2 = 0 \Rightarrow \sigma_x^2 \sigma_y^2 + \lambda^2 - \lambda(\sigma_x^2 + \sigma_y^2) - \sigma_{xy}^2 = 0$$

ricaviamo λ_1 e λ_2 (σ_I e σ_{II}):

$$\sigma_{I,II} = \lambda_{1,2} = \frac{\sigma_x^2 + \sigma_y^2}{2} \pm \frac{1}{2} \sqrt{(\sigma_x^2 + \sigma_y^2)^2 - 4(\sigma_x^2 \sigma_y^2 - \sigma_{xy}^2)} \quad 7.75a$$

cioè:

$$\sigma_{I,II} = \lambda_{1,2} = \frac{\sigma_x^2 + \sigma_y^2}{2} \pm \frac{1}{2} \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2} \quad 7.75b$$

In alternativa, ricavando $\sin 2\alpha$ in funzione di $\operatorname{tg} 2\alpha$ ricavato con la 7.74 si ottiene:

$$\sigma_{I,II} = \lambda_{1,2} = \frac{1}{2} (\sigma_x^2 + \sigma_y^2) \pm \sigma_{xy} / \sin 2\alpha \quad 7.75c$$

L'inclinazione è data dagli autovettori che rappresentano i coseni direttori degli assi principali. Basta sostituire i valori di λ e normalizzare:

$$v = \begin{bmatrix} \sigma_x^2 - \lambda_1 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 - \lambda \end{bmatrix} = 1$$

$$[\sigma_{xy}^2 + (\sigma_y^2 - \lambda_2)^2]^{1/2} = 1$$

7.9 SUCCESSIONI DI VARIABILI CASUALI

Sia $\{\underline{x}_n\}$ una successione di variabili casuali. Si dice che $\{x_n\}$ tende stocasticamente a zero per $n \rightarrow \infty$ se:

$$\lim_{n \rightarrow \infty} P(|\underline{x}_n| < \varepsilon) = 1 \quad \forall \varepsilon > 0$$

Ciò significa che $\{\underline{x}_n\}$ tende alla variabile casuale x concentrata nell'origine ($P(x=0) = 1$).

Usando il teorema di Tchebjeff si può così dimostrare che:

Condizione sufficiente affinché $\{\underline{x}_n\}$ converga stocasticamente a zero è che:

$$\lim_{n \rightarrow \infty} M[\underline{x}_n] = 0 \quad 7.76$$

$$\lim_{n \rightarrow \infty} \sigma^2[\underline{x}_n] = 0 \quad 7.77$$

Diremo poi che converge stocasticamente a $\{\underline{x}_n\}$ se $\{\underline{x} - \underline{x}_n\}$ converge stocasticamente a zero.

7.10 CONVERGENZA «IN LEGGE»

Oltre alla convergenza stocastica della successione di vc $\{\underline{x}_n\}$ ad \underline{x} si può definire una convergenza in legge:

Si dice che $\{\underline{x}_n\}$ tende ad \underline{x} «in legge» se, essendo $\{F_n(x)\}$ la successione delle funzioni di distribuzione di $\{\underline{x}_n\}$ ed $F(x)$ la funzione di distribuzione di \underline{x} si ha:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad 7.78$$

Questo tipo di convergenza serve per studiare il comportamento asintotico di somme di variabili casuali del tipo:

$$S_n = \sum_{i=1}^n x_i \quad \text{per } n \rightarrow \infty \quad 7.79$$

Si può dimostrare infatti che sotto opportune ipotesi sulla successione delle $\{x_i\}$ la successione $\{S_n\}$ tende asintoticamente in legge ad una distribuzione normale.

7.11 TEOREMA CENTRALE DELLA STATISTICA

Teorema

Sia $\{x_i\}$ una successione di variabili casuali indipendenti, tutte con la stessa distribuzione e con:

$$M[x_i] = \mu; \quad \sigma^2(x_i) = \sigma^2$$

Allora la successione:

$$S_n = \sum_{i=1}^n x_i$$

tende asintoticamente in legge (si indica con il simbolo \sim) alla normale del tipo:

$$S_n \sim N[n\mu, n\sigma^2] \quad 7.80$$

\forall distribuzione delle $\{x_i\}$.

Prima osservazione al teorema centrale della statistica

Il teorema interpreta un fatto riconosciuto sperimentalmente – gli errori di misura tendono a distribuirsi normalmente – quando il procedimento di misura è usato al limite della sua precisione massima.

Gli errori di misura cioè dipendono da una serie di fattori ambientali, strumentali e

soggettivi che hanno, ciascuno isolatamente, influenza impercettibile sul procedimento di misura ($\mu \cong 0, \sigma^2 \cong 0$), ciascuno di questi fattori assume anche perciò l'aspetto di una vc indipendente dalle altre (umidità, pressione, temperatura, luminosità ecc.).

Tutti questi fattori assieme producono tuttavia un effetto sensibile: l'errore di misura, che sarà descritto dalla vc somma di molte altre. Per il teorema centrale l'errore di misura tende ad essere distribuito normalmente $N[n\mu, n\sigma^2]$.

Seconda osservazione al teorema centrale della statistica

Il teorema è meno teorico di quanto possa apparire perché permette di usare la normale N come distribuzione approssimata di quantità importanti come il valore medio m (media campionaria).

Sia x una vc comunque distribuita e sia $\{x_1, x_2, \dots, x_n\}$ la vc n -dimensionale generata pensando di ripetere n estrazioni dalla vc x . La $\{x_1, x_2, \dots, x_n\}$ descrive i campioni di numerosità n della x . La media campionaria vale:

$$m = \frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n \frac{x_i}{n} \quad \text{con } \frac{x_i}{n} \text{ v.c. indipendenti}$$

Nell'ipotesi che, per ciascun x_i :

$$M[x_i] = \mu; \quad \sigma^2(x_i) = \sigma^2$$

sarà dunque:

$$M\left[\frac{x_i}{n}\right] = \frac{\mu}{n}; \quad \sigma^2\left(\frac{x_i}{n}\right) = \frac{\sigma^2}{n^2}$$

Se supponiamo che il campione sia numeroso (n grande) possiamo applicare ad m il teorema centrale e dire che \forall distribuzione iniziale di x , m tenderà asintoticamente in legge a:

$$m \sim N\left[n \frac{\mu}{n}, n \frac{\sigma^2}{n^2}\right] = N\left[\mu, \frac{\sigma^2}{n}\right] \tag{7.81}$$

Si noti che, se si volesse ricavare la distribuzione esatta di m cioè di $\frac{1}{n} \sum_{i=1}^n x_i$, si dovrebbero calcolare n integrali di convoluzione seguenti (infatti le x_i sono indipendenti):

$$f(m) = \int_{-\infty}^{\infty} f_{x_1}(x_1) f_{x_2}(x_2) \dots f_{x_n}(x_n) dx_1 \dots dx_n$$

nel caso particolare, siccome $f_{x_i}(x_i) = f_{x_j}(x_j) = f(x)$ si dovrebbero calcolare n integrali di convoluzione di $f(x)$ con se stessa.

È anche matematicamente possibile dimostrare il teorema, infatti, presa una qualsiasi $f(x)$ di partenza, l'integrale di convoluzione di $f(x)$ con se stessa tende, per n grande, alla funzione di Gauss.

Si noti che la 7.81 giustifica il fatto che come valore rappresentativo della popolazione si scelga la media campionaria: rispetto ad una qualsiasi x_i ha varianza n volte minore.

7.12 LE STATISTICHE CAMPIONARIE E I CAMPIONI BERNOULLIANI

Definiamo campione Bernoulliano, tratto da una vc x (che descrive l'esperimento stocastico ξ), l'insieme dei risultati ottenuti dalla ripetizione per n volte in maniera indipendente dello stesso esperimento ξ (esempio: l'estrazione da un'urna con sostituzione).

Osservazione

Lo stesso campione Bernoulliano, per l'indipendenza, può essere visto alternativamente o come risultato di n estrazioni dalla vc x o come estrazione da una vc a n -dimensioni $(x_1 \dots x_n)$ tutte indipendenti e tutte distribuite come x . (Esempio: il lancio di una moneta n volte e il lancio di n monete una sola volta).

Se x ha densità di probabilità $f_x(x)$ la x^n ha densità:

$$f_{x^n} = f_{x^n}(x_1 \dots x_n) = f_x(x_1) f_x(x_2) \dots f_x(x_n) \quad 7.82$$

per l'ipotesi di indipendenza.

Definizione di *statistica campionaria*

La statistica campionaria t è un (\forall) operatore statistico applicato a una variabile campionaria.

Ad esempio:

$$t = t(x_1, x_2, \dots, x_n); \quad 7.83$$

t può essere la media campionaria, la varianza campionaria, il momento di ordine m campionario, la correlazione campionaria, ecc.

Tutto ciò significa che t sarà a sua volta una vc (a una dimensione) funzione della vc n -dimensionale x^n .

Ad esempio se t è l'operatore media m :

$$m = \frac{1}{n} \sum_{i=1}^n x_i = t_0$$

t_0 rappresenta l'estrazione dalla statistica campionaria t .

7.13 LE STATISTICHE «CAMPIONARIE» COME «STIME» DELLE CORRISPONDENTI QUANTITÀ TEORICHE DELLE VARIABILI CASUALI

Qual è il rapporto tra la vc statistica campionaria t , di cui disponiamo di una estrazione t_0 ed il valore teorico (ϑ) del parametro corrispondente a t ? Ad esempio a m_x corrisponde μ_x , ad s^2 corrisponde σ^2 ; quale rapporto esiste fra questi valori? Il rapporto viene detto stima.

Ad esempio si dice che m è stima di μ , od anche s^2 è stima di σ^2 se è corretta e consistente. Vediamo che significano questi aggettivi.

Stima corretta o non deviata

Si dice che la stima è corretta quando la variabile casuale t ammette come media teorica ϑ :

$$M[t] = M[t(x_1 \dots x_n)] = \vartheta \tag{7.84}$$

Stima consistente

Si ha quando per $n \rightarrow \infty$ la corrispondente successione di variabili casuali t_n tende stocasticamente a ϑ , cioè:

$$\lim_{n \rightarrow \infty} t_n = \vartheta \tag{7.85}$$

Per il teorema centrale della statistica ciò è verificato se:

$$\left\{ \begin{array}{l} \lim_{n \rightarrow \infty} M[t_n] = \vartheta \\ \lim_{n \rightarrow \infty} \sigma^2[t_n] = 0 \end{array} \right. \tag{7.86}$$

Stima efficiente

In molti casi esiste più di una stima corretta e consistente di ϑ , allora si cerca quella stima t più concentrata attorno a ϑ cioè una stima efficiente, definita come la stima t di ϑ di minima varianza.

Stima di massima verosimiglianza

Vi è infine la stima di massima verosimiglianza che consiste nel trovare quell'operatore t che rende massima una funzione \mathbf{L} detta di verosimiglianza.

Come esempio ed esercizio vediamo se la media campionaria m può essere presa come stima della quantità teorica μ .

La media campionaria m è una stima corretta e consistente della media teorica μ della vc x , infatti soddisfa a:

– *correttezza*

$$M[m] = M\left[\frac{1}{n} \sum x_i\right] = \frac{1}{n} \sum M[x_i] = \frac{1}{n} n\mu = \mu \quad (\forall i = 1 \dots n)$$

– *consistenza*: per quanto visto la 7.86 è facilmente provata,

$$\lim_{n \rightarrow \infty} M[m] = \mu \quad 7.88$$

Per provare la 7.87 si può scrivere:

$$m = \sum \frac{x_i}{n}$$

Per la propagazione della varianza ricaviamo:

$$\sigma^2(m) = \sum \frac{1}{n^2} \sigma^2(x_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

ed allora è facile vedere che:

$$\lim_{n \rightarrow \infty} \sigma^2(m) = 0 \quad \text{C.V.D.} \quad 7.89$$

Si può verificare che tutte le stime lineari $m' = \sum \lambda_i x_i$ tali che $\sum \lambda_i = 1$ sono stime corrette di μ ma m è quella di minima varianza (cioè efficiente).

Cerchiamo infatti il minimo della quantità:

$$\sigma^2(m') = \sum \lambda_i^2 \sigma^2$$

con la condizione:

$$\sum \lambda_i = 1$$

Questo è un problema di minimo condizionato che si risolve con i moltiplicatori di Lagrange minimizzando la funzione:

$$\phi = \sum \lambda_i^2 \sigma^2 + ((\sum \lambda_i) - 1) \cdot k = \min$$

Il differenziale totale di ϕ dovrà annullarsi:

$$\forall \frac{\partial \phi}{\partial \lambda_i} = 0 \Rightarrow 2\sigma^2 \lambda_i + k = 0 \Rightarrow \lambda_i = -\frac{k}{2\sigma^2}$$

$$\sum \lambda_i = -\frac{nk}{2\sigma^2} = 1 \Rightarrow k = -\frac{2\sigma^2}{n}$$

$$\lambda_i = \frac{1}{n} \Rightarrow m' = m \quad \text{C.V.D.}$$

Dunque si sceglie come valore rappresentativo di tutta la popolazione di misure la media campionaria non solo perché ha varianza n volte minore rispetto alla varianza di ciascun campione, ma anche perché ha la minima varianza.

Come ulteriore esempio vediamo se la varianza campionaria s^2 è una stima di σ^2 :

$$s^2 = \frac{1}{n} \sum (x_i - m)^2 = \frac{\sum v_i^2}{n}$$

dove m è la media campionaria.

Verifichiamo la correttezza, se cioè:

$$M[s^2] = \sigma^2$$

scriviamo s^2 in questo modo:

$$\begin{aligned} s^2 &= \frac{1}{n} \sum [(x_i - \mu) + (\mu - m)]^2 = \\ &= \left[\frac{1}{n} \sum (x_i - \mu)^2 \right] + \left[\frac{2}{n} \sum (x_i - \mu)(\mu - m) \right] + [(\mu - m)^2] = \\ &= \frac{1}{n} \sum (x_i - \mu)^2 + 2(m - \mu)(\mu - m) + (\mu - m)^2 \\ s^2 &= \frac{1}{n} \sum (x_i - \mu)^2 - (m - \mu)^2 \end{aligned} \quad 7.90$$

Applichiamo alla 7.89 l'operatore media:

$$M[s^2] = \frac{1}{n} \sum M[(x_i - \mu)^2] - M[(m - \mu)^2]$$

per definizione:

$$M[(x_i - \mu)^2] = \sigma^2$$

inoltre:

$$\begin{aligned} M[(m - \mu)^2] &= \sigma^2(m) = \frac{\sigma^2}{n} \\ M[s^2] &= \frac{1}{n} \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 \neq \sigma^2 \end{aligned} \quad 7.91$$

Cioè la stima non è corretta. Si dimostra che è invece corretta la stima dell'operatore \bar{s}^2 ($M[\bar{s}^2] = \sigma^2$) definita da:

$$\bar{s}^2 = \frac{\sum (x_i - m)^2}{(n-1)} \quad 7.92$$

ed è consistente; infatti è facile verificare che:

$$\lim_{n \rightarrow \infty} \sigma^2(\bar{s}^2) = \left(\frac{n}{n-1} \right)^2 \sigma^2(\bar{s}^2) = 0$$

7.14 FUNZIONE DI VEROSIMIGLIANZA E PRINCIPIO DI MASSIMA VEROSIMIGLIANZA

Partiamo al solito dalla vc n -dimensionale x descritta dalla funzione $f_x(x_1 \dots x_n)$ secondo la forma 7.82 ma ora anche in funzione di operatori statistici ϑ , ad esempio $\vartheta = [\mu, \sigma^2]^T$, cioè esprimiamo la funzione f attraverso:

$$f_x(x_i, \vartheta)$$

per le ipotesi di indipendenza delle n variabili x_i , ricordando ancora la 7.82:

$$f_{\underline{x}}(x_i, \vartheta) = \prod_{i=1}^n f_x(x_i, \vartheta) = \mathbf{L}(x_i, \vartheta) \quad 7.93$$

Il secondo uguale definisce la funzione \mathbf{L} detta di verosimiglianza (*likely hood*).

È evidente che nulla abbiamo detto sul generico ϑ ; un criterio di scelta è prendere un valore generico t e cercare di rendere massima $\mathbf{L}(x_i, \vartheta)$ verificando che sia massima per $\vartheta=t$, cioè cercare:

$$\exists t / \max_{\vartheta=t} \mathbf{L}(x_i, \vartheta) \Rightarrow \frac{\partial \mathbf{L}}{\partial \vartheta} = 0 \Rightarrow \frac{\partial \log(\mathbf{L})}{\partial \vartheta} = 0 \quad 7.94$$

cioè, per la 7.93:

$$\sum_{i=1}^n \frac{\partial f(x_i, \vartheta)}{\partial \vartheta} = 0 \quad 7.95$$

Ad esempio per la variabile normale standardizzata z^n :

$$\mathbf{L} = f_{\underline{x}} = \prod f_x(x_i) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum (x_i - \mu_i)^2}{2\sigma^2}} \quad 7.96$$

si ha in questo caso:

$$\vartheta = [\mu, \sigma^2]^T$$

Il valore massimo di \mathbf{L} si ha cercando il minimo dell'esponente:

$$+ \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_i)^2 = \frac{1}{2\sigma^2} \mathbf{v}^T \mathbf{v} = \min \quad 7.97$$

con ϑ , variabile scarto. In questo caso il principio di massima verosimiglianza porta alla stima di minima varianza e cioè alla ricerca di uno stimatore efficiente.

Per variabili normali non standardizzate, ricordando la 7.67 e la 7.69 occorre rendere minima la quantità:

$$\chi_n^2 = (x - \mu)^T C_{xx}^{-1} (x - \mu) = \mathbf{v}^T C_{xx}^{-1} \mathbf{v} = \min \quad 7.98$$

La 7.98 spesso viene scritta utilizzando un'altra matrice definita matrice dei pesi P , (σ_0^2 è una costante positiva):

$$P = C_{xx}^{-1} \sigma_0^2 \quad 7.99$$

È questo il principio dei minimi quadrati che, nel caso in cui P sia una matrice diagonale, può essere scritto nella forma:

$$\sum_{i=1}^n p_i \mathbf{v}_i^2 = \min = \sigma_0^2 \chi_n^2 \quad 7.100$$

Dobbiamo tuttavia affermare che la stima di minima varianza, che coincide con quella di massima verosimiglianza per variabili normali e che porta al principio dei minimi quadrati, prescinde da ipotesi sulla distribuzione delle misure.

7.15 LA MEDIA PONDERATA (O PESATA)

Poniamo di eseguire n misure di una v.c x , fatte con diversa precisione ma indipendenti tra loro; ciascuna x_i può considerarsi come estrazione da popolazioni con diverse varianze $\sigma^2(x_i) = \sigma_i^2$ ma con la stessa media μ_x . Ci si chiede quale è la stima più attendibile del valore medio di x . Avevamo verificato per la media campionaria che tutte le stime del tipo:

$$\bar{x} = \sum \lambda_i x_i \tag{7.101}$$

sono corrette, d'altra parte non possiamo usare i valori $\lambda_i = 1/n$ perché il risultato non sarà stima di minima varianza; dovremmo, intuitivamente, pesare di più le x_i con σ_i minore.

Anche qui cerchiamo uno stimatore \bar{x} che sia stima efficiente, e troviamo il minimo condizionato attraverso i moltiplicatori di Lagrange:

$$\left\{ \begin{array}{l} \sigma^2(\bar{x}) = \sum \lambda_i^2 \sigma_i^2 = \min \\ \sum \lambda_i = 1 \end{array} \right. \tag{7.102}$$

e minimizziamo la funzione:

$$\begin{aligned} \phi &= \sum \lambda_i^2 \sigma_i^2 - k[\sum \lambda_i - 1] \\ \frac{\partial \phi}{\partial \lambda_i} &= 0 \Rightarrow 2\lambda_i \sigma_i^2 - k = 0 \end{aligned}$$

ricaviamo:

$$\lambda_i = \frac{k}{2} \frac{1}{\sigma_i^2} \tag{7.103}$$

Come per la 7.99, presa una seconda costante positiva, σ_0^2 , viene definito peso il valore:

$$p_i = \frac{\sigma_0^2}{\sigma_i^2} \tag{7.104}$$

cosicché la 7.103 può scriversi:

$$\lambda_i = \frac{k}{2} \frac{p_i}{\sigma_0^2}$$

ma, imponendo la seconda delle 7.102, si ricava k :

$$k = \frac{2\sigma_0^2}{\sum p_i}$$

per cui la 7.103 può essere riscritta:

$$\lambda_i = \frac{\sigma_0^2}{\sum p_i \sigma_0^2} \frac{p_i}{\sigma_0^2} = \frac{p_i}{\sum p_i} \quad 7.105$$

dunque la 7.101 diviene:

$$\bar{x} = \frac{\sum p_i x_i}{\sum p_i} \quad 7.106$$

Si nota pure che il minimo cercato nella stima di $\sigma^2(x)$ vale:

$$\sigma^2(\bar{x}) = \min = \sum \lambda_i \sigma_i^2 = \frac{\sum p_i^2 \sigma_i^2}{(\sum p_i)^2} \quad 7.107$$

Se non si conoscono i valori σ_i^2 ma si conoscono solo i pesi p_i e \bar{x} (dalla 7.106), la 7.107 non è direttamente utilizzabile. Dopo il calcolo di \bar{x} , si dimostra che:

$$\sigma_0^2 = \frac{1}{n-1} \sum p_i (x_i - \bar{x})^2 = \frac{\sum p_i v_i^2}{n-1} \quad 7.108$$

$$\sigma^2(\bar{x}) = \sigma_0^2 \frac{1}{\sum p_i} = \frac{\sum p_i v_i^2}{(n-1) \sum p_i} \quad 7.109$$

8. APPLICAZIONI DEL PRINCIPIO DEI MINIMI QUADRATI AL TRATTAMENTO DELLE OSSERVAZIONI

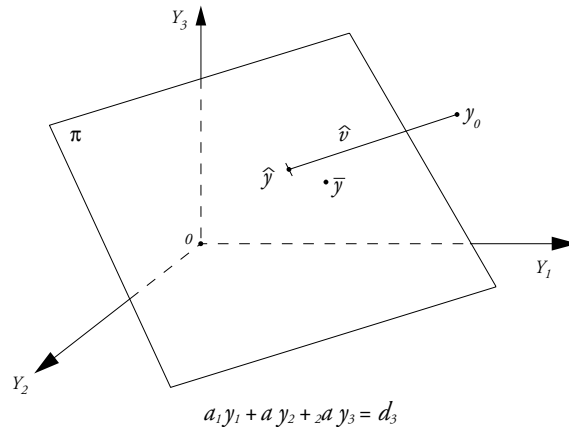


Fig. 8.1

Prendiamo in esame la variabile casuale tridimensionale $\underline{y} = (y_1, y_2, y_3)$ che rappresenta le misure che possono essere fatte su un esperimento \mathcal{E} del quale si conosca già un modello fisico, lineare del tipo:

$$a_1 y_1 + a_2 y_2 + a_3 y_3 = d \quad 8.1$$

che rappresenta l'equazione di un piano nello spazio detto piano delle misure ammissibili. Sia ad esempio \mathcal{E} l'esperimento la misura dei tre angoli di un triangolo piano: la somma di questi deve essere uguale a π .

Facciamo poi l'ipotesi che le misure \underline{y} abbiano distribuzione normale (ipotesi non indispensabile), media diversa da zero e varianza unitaria, vale a dire:

$$\underline{y} = N[\bar{y}, I] \quad 8.2$$

Queste ipotesi vengono definite modello stocastico.

Della variabile casuale y si conosce una estrazione, la misura y_0 che, a causa della dispersione di y non è detto soddisfi la 8.1. A causa di errori accidentali infatti y_0 è fuori da questo piano ad una distanza \hat{v} . In genere cioè si ha:

$$a_1 y_{01} + a_2 y_{02} + a_3 y_{03} - d = u \neq 0 \quad 8.3$$

tuttavia, siccome y_0 è estratto dalla stessa variabile casuale y , il suo valore medio sarà identico al valore medio di y :

$$M[y_0] = \bar{y} \quad 8.4$$

Ora noi cerchiamo una stima \hat{y} di y (il simbolo \hat{y} sta per stima di massima verosimiglianza) che sia la più vicina possibile a y_0 ma che appartenga ancora ai valori ammissibili del piano π ; in questo caso è intuitivo scegliere per \hat{y} la normale a π condotta da y_0 , cioè:

$$\hat{y} = y_0 - \hat{v} \quad 8.5$$

tale che renda minimo lo scalare distanza al quadrato:

$$d^2 = \hat{v}^T \hat{v} = (y_0 - \hat{y})^T (y_0 - \hat{y}) = \min \quad 8.6$$

Vedremo ora se questa equazione è sufficiente a risolvere il problema, si tratta cioè di ricavare \hat{y} e le caratteristiche della dispersione di \hat{y} a partire dalle ipotesi stocastiche su y 8.2, dal modello geometrico 8.1 e dalle condizioni di stima 8.6. Nel caso in cui $C_{yy} \neq I$ la 8.6 si modifica nella già nota equazione di minimi quadrati:

$$d^2 = (y_0 - \hat{y})^T C_{yy}^{-1} (y_0 - \hat{y}) = \min \quad 8.7$$

Il principio dei minimi quadrati, che coincide con il principio di massima verosimiglianza nel caso di distribuzione normale, conduce a trovare uno stimatore efficiente di minima norma: la distanza al quadrato 8.7 si chiama infatti norma quadratica del vettore \hat{v} . Il minimo di detta norma rimane tale, come dimostrato e come è ovvio, anche per trasformazioni lineari del sistema di riferimento. Che la 8.7 esprima poi una distanza è evidente; partendo infatti da variabili casuali x con $C_{xx} = I$, il minimo della distanza quadratica vale appunto:

$$(x_0 - \hat{x})^T - (x_0 - \hat{x}) = \min \quad 8.8$$

scegliendo una qualsiasi matrice di rotazione per cui:

$$(y_0 - \hat{y}) = R^{-1}(x_0 - \hat{x})$$

si arriva alla:

$$d^2 = \min = (y_0 - \hat{y})^T R^T R^{-1} (y_0 - \hat{y}) \quad 8.9$$

che è appunto un altro modo di vedere la formula 8.7.

8.1 I MINIMI QUADRATI APPLICATI AD EQUAZIONI DI CONDIZIONE CON MODELLO LINEARE

Vediamo se esiste una soluzione \hat{y} all'equazione 8.7. Per le ipotesi di minimo la soluzione cercata sarà la stessa a meno di una costante moltiplicativa $1/\sigma_0^2$.

È possibile allora cercare questo minimo anche partendo dalla conoscenza della matrice P che, a meno di una costante moltiplicativa σ_0^2 è proporzionale a C_{yy} :

$$\frac{1}{\sigma_0^2} C_{yy} = Q = P^{-1} \Rightarrow P = \sigma_0^2 C_{yy}^{-1} \quad 8.10$$

P è definita matrice dei pesi. Si cerca ora il minimo della quantità scalare:

$$(y_0 - \hat{y})P(y_0 - \hat{y}) = \min \quad 8.11$$

col modello stocastico definito da:

$$P = Q^{-1} = \sigma_0^2 C_{yy}^{-1} = \sigma_0^2 \text{diag}(\sigma_{yi}^2)^{-1} \quad 8.12$$

e le $\ell \leq m$ equazioni di condizione, generalizzazione delle 8.1:

$$D\hat{y} = d^1 \quad 8.13$$

Si desidera ricavare la stima delle quantità:

$$\begin{cases} \hat{y} \approx y \\ \hat{\sigma}_0^2 \approx \sigma_0^2 \\ C_{\hat{y}\hat{y}} \approx C_{yy} \end{cases}$$

(Il simbolo \approx indica: stima di). Per la ricerca del minimo condizionato si utilizzano i moltiplicatori di Lagrange prendendo come funzione obiettivo la funzione Φ costruita con le 8.11 e 8.13:

$$\Phi = \frac{1}{2}(y_0 - \hat{y})^T P(y_0 - \hat{y}) + (D\hat{y} - d)\lambda \quad 8.14$$

con:

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_\ell) \quad \ell \leq m \quad 8.15$$

dove ℓ è il numero di condizioni ed m il numero di misure.

Imponendo la stazionarietà della funzione Φ si ha:

$$d\Phi = 0 = -d\hat{y}^T P(y_0 - \hat{y}) + d\hat{y}^T D^T \lambda = 0 \quad \forall d\hat{y}^T$$

¹ Ad esempio, nel caso della misura di angoli interni di un triangolo, si provi a risolvere come segue il problema nell'ipotesi di avere le misure:

$y_0^1 = 60 \text{ gon}; y_0^2 = 70 \text{ gon}; y_0^3 = 70.003 \text{ gon}; \text{con } \sigma_y^2 = \text{cost} = \pm 10^{-3} \text{ gon}.$

cioè:

$$P(y_0 - \hat{y}) = D^T \lambda \quad 8.16$$

$$\hat{y} = y_0 - P^{-1} D^T \lambda \quad 8.17$$

che posta nella 8.13 permette di ricavare:

$$D(y_0 - P^{-1} D \lambda) - d = 0$$

$$Dy_0 - DP^{-1} D^T \lambda - d = 0$$

Se definiamo le matrici:

$$K = DP^{-1} D^T \quad 8.18$$

ed il vettore:

$$U = Dy_0 - d \quad 8.19$$

dove U sono definiti errori di chiusura, si ha:

$$K \lambda = (Dy_0 - d); \quad \lambda = K^{-1}(Dy_0 - d)$$

Quest'ultima, posta nella 8.17 permette di ricavare \hat{y} :

$$\hat{y} = y_0 - P^{-1} D^T K^{-1} U \quad 8.20$$

Si dimostra poi che la stima di σ_0^2 vale:

$$\hat{\sigma}_0^2 = \frac{U^T K^{-1} U}{\ell} \quad 8.21$$

Esempio applicativo: anello di livellazione

Si sono misurati i tre dislivelli Δ_{12}^0 ; Δ_{23}^0 ; Δ_{13}^0 , di un anello di tre lati, attraverso una livellazione geometrica. Si sa che in questo caso:

$$\sigma_{\Delta} = \pm \alpha \sqrt{D}$$

dove α è una costante e D è la distanza percorsa fra i punti espressa in km. Si sa che i dislivelli debbono soddisfare all'equazione:

$$\Delta_{12} + \Delta_{23} - \Delta_{13} = 0$$

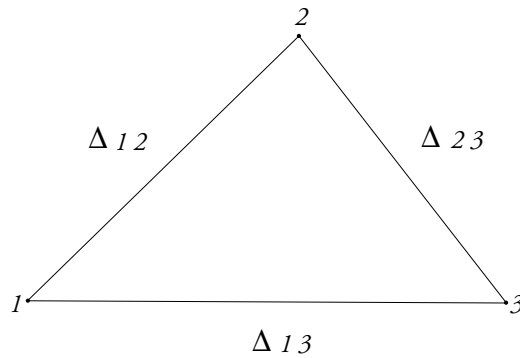


Fig. 8.2 – Anello di livellazione.

Applicando le formule risolutive ricavate nell'esempio proposto si ha:

$$\underline{\underline{D}} = (1 \quad 1 \quad -1)$$

$$C_{yy}^{-1} = \sigma_0^2 \cdot \text{diag}\left(\frac{1}{D_{12}}; \frac{1}{D_{23}}; \frac{1}{D_{31}}\right)$$

Per semplicità, nel calcolo della matrice dei pesi, possiamo trascurare la costante σ_0^2 e porre:

$$Q = P^{-1} = \text{diag}(D_{12}; D_{23}; D_{31})$$

Applicando la 8.18 si ha:

$$K = (1, 1, -1)\text{diag}(D_{12}; D_{23}; D_{31})(1, 1, -1)^T$$

e cioè in definitiva:

$$K = \sum D_{ij}$$

$$U = Dy_0 - d = (\Delta_{12}^0 + \Delta_{23}^0 - \Delta_{13}^0)$$

ed applicando la 8.20 si ha:

$$\hat{y} = (\Delta_{12}^{\circ}, \Delta_{23}^{\circ}, \Delta_{13}^{\circ})^T - \text{diag}(D_{12}, D_{23}, D_{31}) \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} K^{-1} (\Delta_{12}^{\circ} + \Delta_{23}^{\circ} - \Delta_{13}^{\circ})$$

Si ricava infine la soluzione:

$$\hat{y}_1 = \hat{\Delta}_{12} = \Delta_{12}^{\circ} - \frac{D_{12}U}{\sum D_{ij}}$$

$$\hat{y}_2 = \hat{\Delta}_{23} = \Delta_{23}^{\circ} - \frac{D_{23}U}{\sum D_{ij}}$$

$$\hat{y}_3 = \hat{\Delta}_{13} = \Delta_{13}^\circ - \frac{D_{13}U}{\sum D_{ij}}$$

Si ottiene dunque quanto intuitivamente si poteva già capire: che cioè l'errore di chiusura U si ripartisce in tre parti, proporzionali secondo la formula della media ponderata, con pesi D_{ij} , che sono le distanze fra i capisaldi altimetrici delle reti.

8.2 MINIMI QUADRATI, FORMULE RISOLUTIVE NEL CASO DELL'UTILIZZO DI PARAMETRI AGGIUNTIVI

Sia dato un modello stocastico definito dai valori osservati (campione m -dimensionale):

$$y_0 = \begin{pmatrix} y_{01} \\ \vdots \\ y_{0m} \end{pmatrix} \quad \text{tratto da} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

che ipotizziamo abbia media:

$$M[y] = \bar{y} \tag{8.22a}$$

e dispersione:

$$C_{yy} = \sigma_0^2 Q = \sigma_0^2 P^{-1} \tag{8.22b}$$

con σ_0^2 costante positiva incognita e Q (o P) matrice nota e definita positiva.

Per ipotesi il modello deterministico è ancora lineare.

Per motivi fisici o geometrici ipotizziamo che y sia ristretto a stare su un iperpiano π (varietà lineare) a n -dimensioni con $n < m$, del tipo:

$$\underline{y} = \underline{A} \underline{x} + \underline{a} \tag{8.23}$$

con $r(A) = n$, vale a dire $A^T A$ risulta di rango n_1 pieno e invertibile.

Le dimensioni di \underline{x} ed \underline{a} , (che brevemente in seguito indicheremo senza sottolineature) sono:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}; \quad a = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \tag{8.24}$$

Le componenti di x sono dette parametri aggiuntivi, o più spesso solo parametri.

In funzione delle misure y_0 estratte da y si vogliono trovare le stime \hat{y} e $\hat{\sigma}_0^2$ di minima varianza:

$$\begin{cases} \hat{y} \approx y \\ \hat{\sigma}_0^2 \approx \sigma_0^2 \end{cases}$$

e la relativa matrice di varianza covarianza $C_{\hat{y}\hat{y}}$. Occorre trovare, ricordando la 8.5:

$$\min (y_0 - \hat{y})^T P (y_0 - \hat{y}) = \min \mathbf{v}^T P \mathbf{v} = \min \chi_{m-n}^2 \quad 8.25$$

con le $(m-n)$ condizioni aggiuntive: $\hat{y} \in \pi$, cioè:

$$\hat{y} = A \hat{x} + a \quad 8.26$$

Come altrove si è notato, la 8.25 rappresenta il minimo di una distanza generalizzata secondo la «metrica» P , mentre la 8.26 esprime il fatto che alle variabili casuali y sono legati n parametri aggiuntivi \hat{x} che dipendono nel modo lineare 8.26 dalle misure \hat{y} .

Anche qui il problema si risolve con i moltiplicatori di Lagrange, si cerca il minimo condizionato della funzione $\Phi(\hat{x}, \hat{y})$:

$$\Phi(\hat{x}, \hat{y}) = \frac{1}{2} (y_0 - \hat{y})^T P (y_0 - \hat{y}) + (\hat{y} - A \hat{x} - a) \lambda = \min \quad 8.27$$

con:

$$\lambda = (\lambda_1 \dots \lambda_m)^T; \quad n < m \quad 8.28$$

si ha:

$$d\Phi = -d\hat{y}^T P (y_0 - \hat{y}) + d\hat{x}^T \lambda - d\hat{x}^T A^T \lambda = 0$$

Annullando i termini che moltiplicano i due differenziali si devono soddisfare le equazioni:

$$A^T \lambda = 0 \quad 8.29$$

$$-P(y_0 - \hat{y}) + \lambda = 0 \quad 8.30$$

Da quest'ultima si ottiene, essendo la matrice P definita positiva:

$$\hat{y} = y_0 - P^{-1} \lambda$$

ma, ricordando anche la 8.26:

$$P^{-1} \lambda = y_0 - A \hat{x} - a$$

allora:

$$\lambda = P(y_0 - a) - PA \hat{x}$$

che, sostituita nella 8.29 permette di scrivere:

$$A^T P (y_0 - a) - A^T P A \hat{x} = 0$$

cioè:

$$\hat{x} = (A^T P A)^{-1} A^T P (y_0 - a) \quad 8.31a$$

e, dalla 8.26 si può ricavare \hat{y} .

Definito poi vettore dei termini noti l :

$$(y_0 - a) = l \quad 8.32$$

e definita matrice normale N :

$$N = A^T P A \quad 8.33$$

si può anche scrivere:

$$\hat{x} = N^{-1} A^T P l \quad 8.31b$$

Infine si può dimostrare che la stima $\hat{\sigma}_0^2$ di σ_0^2 vale:

$$\hat{\sigma}_0^2 = \frac{(y_0 - \hat{y})^T P (y_0 - \hat{y})}{m - n} = \frac{\hat{v}^T P \hat{v}}{m - n} \quad 8.34$$

dove il numero intero:

$$r = m - n \quad 8.35$$

viene detta ridondanza globale o ridondanza.

Lo scalare $\hat{\sigma}_0^2$, (a parte la costante r), rappresenta dunque la distanza quadratica del vettore \hat{v} nella metrica P o, in alternativa, il valore χ^2 della 8.25.

Dalla 8.32 e dalla definizione di \hat{v} ricaviamo:

$$\hat{v} = l - A \hat{x} \quad 8.36$$

Si dimostra che la matrice di varianza covarianza dei parametri compensati vale:

$$C_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 N^{-1}; \quad 8.37$$

È possibile ricavare inoltre la matrice di varianza covarianza degli scarti, dopo la compensazione:

$$C_{\hat{v}\hat{v}} = \hat{\sigma}_0^2 [P^{-1} - A N^{-1} A^T] \quad 8.38$$

Infine si può dimostrare che la matrice:

$$R = \frac{1}{\hat{\sigma}_0^2} P C_{\hat{v}\hat{v}} \quad 8.39$$

$$R = I - P A N^{-1} A^T \quad 8.40$$

è una matrice di dimensione $m \bullet m$ detta di ridondanza, contenente dei numeri puri ed indipendente dal sistema di riferimento scelto. La proprietà di questa matrice è indicare il contributo che ogni singola misura apporta alla ridondanza globale $r = m - n$. Si può dimostrare infatti che:

$$\text{tr}(R) = \sum_l^m r_{jj} = (m - n) = r \quad 8.41$$

con r_{jj} chiamato ridondanza locale dell'osservazione j . Osservando la 8.41 si nota che è possibile ricavare R senza aver eseguito le misure y_0 .

Similmente possiamo notare che anche altre formule già ricavate non dipendono dalle misure eseguite.

Più in generale, nel caso in cui il problema sia il progetto di una rete topografica, si possono ricavare a priori le precisioni dei parametri, la precisione delle misure dopo la compensazione, il contributo delle stesse alla rigidità della rete. È cioè possibile già in fase di progetto della rete prevedere le precisioni finali, togliere le misure poco significative, o che potrebbero nascondere errori che più facilmente sfuggono ai test di controllo, migliorare infine l'affidabilità della rete.

Esempio applicativo

Compensiamo secondo il metodo dei parametri aggiuntivi la rete di livellazione precedentemente vista (fig. 8.2). Si sono misurati i dislivelli:

$$\begin{aligned} \Delta_{12} &= Q_2 - Q_1 \\ \Delta_{23} &= Q_3 - Q_2 \\ \Delta_{13} &= Q_3 - Q_1 \end{aligned}$$

Si possono identificare i vettori:

$$y = \begin{Bmatrix} \Delta_{12} \\ \Delta_{23} \\ \Delta_{13} \end{Bmatrix}; x = \begin{Bmatrix} Q_2 \\ Q_3 \end{Bmatrix}; \hat{y} = \begin{Bmatrix} \hat{\Delta}_{12} \\ \hat{\Delta}_{23} \\ \hat{\Delta}_{13} \end{Bmatrix}; y_0 = \begin{Bmatrix} \Delta_{12}^0 \\ \Delta_{23}^0 \\ \Delta_{13}^0 \end{Bmatrix}$$

Dei dislivelli, che possono ritenersi misurati in modo indipendente, si conoscono i valori misurati $\Delta_{12}^0, \Delta_{23}^0, \Delta_{13}^0$ con livellazione geometrica. Si conosce anche:

$$\sigma_{\Delta_{ij}} = \pm 1 \text{ mm} \sqrt{D}$$

ove D è la distanza fra i punti espressa in km; per queste ipotesi si potrà porre:

$$P = Q^{-1} = \text{diag}(\sigma_{\Delta_{12}}^2, \sigma_{\Delta_{23}}^2, \sigma_{\Delta_{13}}^2)^{-1}$$

che in questo caso diviene:

$$P = (1 \text{ mm})^{-2} \text{diag}\left(\frac{1}{D_{12}}, \frac{1}{D_{23}}, \frac{1}{D_{31}}\right)$$

I parametri incogniti sono le quote dei tre vertici.

Se effettivamente decidessimo di mantenere come parametri incogniti tutte queste tre quote troveremmo tuttavia ben presto una deficienza di rango nella matrice normale N . A cosa è dovuta? Nel passaggio dallo «spazio delle misure» allo «spazio dei parametri» dobbiamo considerare in questo caso che le prime, essendo invarianti per traslazione, sono definite a meno di una traslazione del sistema di riferimento.

Nel caso dell'utilizzo dei parametri aggiuntivi «coordinate» in un problema ai

minimi quadrati, occorre allora definire (anche arbitrariamente) questo sistema di riferimento, detto *datum*, dal quale non dipendono le misure ma dipendono invece i parametri aggiuntivi. Nel caso in esame ciò si fa, senza perdere di generalità, fissando ad esempio la quota del punto 1 (ad esempio $Q_1 = 0$ m). In tal modo rimangono incognite solo le quote dei punti 2 e 3.

Nell'esempio proposto si ha $n = 2$ (numero dei parametri incogniti) ed $m = 3$ (numero di misure) per cui $r = 1$. La relazione 8.26 si scrive:

$$\hat{y} = \begin{Bmatrix} \hat{\Delta}_{12} \\ \hat{\Delta}_{23} \\ \hat{\Delta}_{13} \end{Bmatrix} = \begin{Bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{Bmatrix} \begin{Bmatrix} \hat{Q}_2 \\ \hat{Q}_3 \end{Bmatrix} + \begin{Bmatrix} -Q_1 \\ 0 \\ -Q_1 \end{Bmatrix} = \underline{\underline{A}} \hat{x} + a$$

Si ha poi:

$$(y_0 - a) = \ell = \begin{pmatrix} \Delta_{12}^0 & + Q_1 \\ & \Delta_{23}^0 \\ \Delta_{13}^0 & + Q_1 \end{pmatrix}$$

La matrice normale vale:

$$A^T P A = N = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \text{diag}(D_{ij})^{-1} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{pmatrix}$$

Sviluppando i calcoli si ottiene:

$$n_1^1 = \frac{1}{D_{12}} + \frac{1}{D_{23}}$$

$$n_{12} = n_{21} = -\frac{1}{D_{23}}$$

$$n_2^2 = \frac{1}{D_{23}} + \frac{1}{D_{13}}$$

ed il vettore b , formato da due valori, risulta:

$$A^T P l = b$$

$$b_1^1 = \frac{1}{D_{12}}(\Delta_{12} + Q_1) - \frac{1}{D_{23}}\Delta_{23}$$

$$b_2^1 = \frac{1}{D_{23}}\Delta_{23} + \frac{1}{D_{13}}\Delta_{13}$$

Ora si può risolvere il sistema od invertire la matrice N e ricavare:

$$\hat{x} = N^{-1}b$$

Si verifica inoltre, (numericamente è in questo caso più facile), che la stima delle misure:

$$\hat{y} = A(N^{-1}b) + a$$

è la stessa ricavata con il metodo delle sole equazioni di condizione, visto per lo stesso esempio.

Si ricavano infine gli scarti:

$$\hat{v} = y_0 - \hat{y} = \Delta_{ij}^0 - \hat{\Delta}_{ij}$$

che permettono di calcolare la 8.34:

$$\sigma_0^2 = \frac{\hat{v}^T P \hat{v}}{3-2} = \min$$

che, ancora, deve risultare identico al valore calcolabile con la 8.21.

8.3 MINIMI QUADRATI: EQUAZIONI DI CONDIZIONE E PARAMETRI AGGIUNTIVI

È questo il caso misto che comprende i due precedentemente trattati.

Premettiamo subito che è difficile poter applicare i risultati che si otterranno in questo caso al calcolo automatico, a causa della quasi impossibile generalizzazione del problema per scopi topografici; in programmazione questi problemi si risolvono secondo l'analisi ed i metodi risolutivi visti nel caso delle equazioni ai parametri in quanto è più facile invece ricondurre questo caso al precedente. Daremo tuttavia, per completezza, uno sguardo alla soluzione teorica del problema.

Sia:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

il vettore dei parametri, funzione (lineare) delle m quantità osservate y , (ad esempio le quote sono funzioni lineari dei dislivelli).

Le osservabili y sono legate da ℓ relazioni lineari contenenti n parametri aggiuntivi x ($n < \ell \leq m$) secondo il modello:

$$Dy = Ax + d \tag{8.42}$$

dove le dimensioni coinvolte sono:

$$\begin{pmatrix} m \\ \ell \end{pmatrix} \begin{pmatrix} 1 \\ m \end{pmatrix} = \begin{pmatrix} n \\ \ell \end{pmatrix} \begin{pmatrix} 1 \\ n \end{pmatrix} \begin{pmatrix} 1 \\ \ell \end{pmatrix}$$

Al solito le ipotesi stocastiche su y sono:

$$y = N[\bar{y}, \sigma_0^2 Q] \text{ con } M[y_0] = \bar{y},$$

Il problema è ricavare le stime:

$$\begin{cases} \hat{y} \approx y \\ \hat{x} \approx x \\ \hat{\sigma}_0^2 \approx \sigma_0^2 \\ C_{\hat{y}\hat{y}} \approx C_{yy} \\ C_{\hat{x}\hat{x}} \end{cases} \quad 8.43$$

secondo il modello fisico²:

$$Dy = Ax + d; \quad r(D) = l; \quad r(A) = n \quad 8.44$$

e infine secondo la condizione di stima:

$$(y_0 - \hat{y})P(y_0 - \hat{y}) = \min$$

Introduciamo i moltiplicatori $\lambda = (\lambda_1 \dots \lambda_n)$ e minimizziamo la funzione:

$$\Phi(\hat{x}; \hat{y}) = (y_0 - \hat{y})^T P(y_0 - \hat{y}) + (D\hat{y} - A\hat{x} - d)^T \lambda \quad 8.45$$

ricaviamo il differenziale:

$$d\Phi = -2d\hat{y}^T P(y_0 - \hat{y}) + d\hat{y}^T D^T \lambda - d\hat{x}^T A^T \lambda = 0$$

che permette ancora di scrivere:

$$\begin{cases} -P(y_0 - \hat{y}) + D^T \lambda / 2 = 0 \\ A^T \lambda = 0 \end{cases}$$

$$\hat{y} = y_0 - P^{-1} D^T \lambda / 2 \quad 8.46$$

ma, ricordando la si ha:

$$Dy_0 - DP^{-1}D^T \lambda / 2 = A\hat{x} + d$$

e ricordando la definizione 8.18 di K:

$$\lambda = 2K^{-1}(Dy_0 - d) - 2K^{-1}A\hat{x} \quad 8.47$$

che inserita nella $A^T \lambda = 0$ ottiene:

$$A^T K^{-1}(Dy_0 - d) - A^T K^{-1}A\hat{x} = 0$$

ed allora:

$$\hat{x} = (A^T K^{-1}A)^{-1} A^T K^{-1}(Dy_0 - d) \quad 8.48a$$

² Intendendo per $r(\bullet)$ il rango del contenuto (\bullet) .

simile alla 8.31a. Chiamando infatti:

$$\begin{aligned} N &= A^T K^{-1} A \\ \hat{x} &= N^{-1} A^T K^{-1} (Dy_0 - d) \end{aligned} \quad 8.48b$$

Ricavato poi λ dalla 8.47:

$$\lambda = 2 K^{-1} (Dy_0 - d - A \hat{x}) = 2 K^{-1} U$$

per definizione di U:

$$U = Dy_0 - A \hat{x} - d \quad 8.49$$

si ha, usando la 8.46:

$$\hat{y} = y_0 - P^{-1} D^T K^{-1} U \quad 8.50$$

8.4 PROPRIETÀ DELLE STIME \hat{y} ED \hat{x} , LORO DISPERSIONE

Le stime \hat{y} ed \hat{x} sono stime corrette di x ed y . Vediamo dapprima la \hat{x} ; ricordiamo che:

$$D\hat{y} = A\hat{x} + d$$

Consideriamo il valore medio di U e ricordiamo la 8.49 e la 8.45:

$$M[U] = Dy - d - Ax = 0$$

ed allora si ha:

$$M[\hat{x}] = (A^T K^{-1} A)^{-1} A^T K^{-1} (DM[y_0] - d) = (A^T K^{-1} A)^{-1} A^T K^{-1} Ax = 0 \quad \text{CVD}$$

Per la correttezza di \hat{y} partiamo considerando la 8.50:

$$M[\hat{y}] = M[y_0] - P^{-1} D K^{-1} M[U] = M[y_0] = y \quad \text{CVD} \quad 8.51$$

Non si dimostra qui la consistenza, si ricorda invece che l'efficienza è l'ipotesi, con la quale ricavammo dette stime e dunque è già verificata.

Cerchiamo ora le matrici di varianza-covarianza delle stime. Chiamiamo con u il vettore:

$$u = Dy_0 - d \quad 8.52$$

la 8.48b assume la forma:

$$\hat{x} = (A^T K^{-1} A)^{-1} A^T K^{-1} u = Su \quad 8.53$$

Propagando la varianza attraverso la 8.52 si ha:

$$C_{uu} = DC_{y_0 y_0} D^T = \sigma_0^2 D Q D^T = \sigma_0^2 K$$

ed ancora propagando, usando stavolta la 8.53:

$$C_{\hat{x}\hat{x}} = SC_{uu}S^T = (A^T K^{-1}A)^{-1} A^T K^{-1} \sigma_0^2 K K^{-1} A (A^T K^{-1}A)^{-1}$$

che semplificata ottiene:

$$C_{\hat{x}\hat{x}} = \sigma_0^2 (A^T K^{-1}A)^{-1} = \sigma_0^2 N^{-1} \quad 8.54$$

Per ottenere la matrice si propaga la varianza a partire dalla 8.50; non si svolge qui il calcolo abbastanza laborioso che permette di ricavare:

$$C_{\hat{y}\hat{y}} = \sigma_0^2 \{Q - QD^T K^{-1} [K - AN^{-1}A^T] K^{-1} DQ\} \quad 8.55$$

con:

$$\sigma_0^2 = \frac{U^T K^{-1}U}{l-n} \quad 8.56$$

dove l è il numero di condizioni o vincoli, n è il numero di parametri incogniti ed m è il numero di misure.

Riassumiamo qui le formule utilizzate nel caso di pure equazioni di condizione e di pure equazioni parametriche utilizzando il risultato generale appena ricavato.

Pure equazioni di condizione

$$D = B; \quad A = 0; \quad K = BQB^T; \quad U = By_0 - b$$

$$\hat{y} = y_0 - QB^T(BQB^T)^{-1}(By_0 - b)$$

$$C_{\hat{u}\hat{u}} = \hat{\sigma}_0^2 K = \hat{\sigma}_0^2 BQB^T$$

$$\hat{\sigma}_0^2 = \frac{(By_0 - b)^T K^{-1}(By_0 - b)}{l - n}$$

$$C_{\hat{v}\hat{v}} = \hat{\sigma}_0^2 QB^T K^{-1} BQ$$

$$C_{\hat{y}\hat{y}} = \hat{\sigma}_0^2 [Q - C_{\hat{v}\hat{v}}]$$

Pure equazioni parametriche

$$D = I; \quad K = Q = P^{-1}; \quad N = A^T P A$$

$$U = (y_0 - \hat{y}) = \hat{v};$$

$$\hat{x} = (A^T P A)^{-1} A^T P (y_0 - d); \quad l = y_0 - d$$

$$\hat{y} = A \hat{x} + d$$

$$\hat{\mathbf{v}} = \mathbf{l} - \mathbf{A}\hat{\mathbf{x}}$$

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T \mathbf{P} \mathbf{v}}{m - n}$$

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = \hat{\sigma}_0^2 \mathbf{N}^{-1};$$

$$\mathbf{C}_{uu} = \mathbf{C}_{\hat{\mathbf{v}}\hat{\mathbf{v}}} = \hat{\sigma}_0^2 [\mathbf{P}^{-1} - \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T]$$

$$\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} = \hat{\sigma}_0^2 \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T$$

Si noti che:

$$\mathbf{C}_{\hat{\mathbf{v}}\hat{\mathbf{v}}} = \mathbf{C}_{yy} - \mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} \quad 8.57$$

Infine la matrice di ridondanza vale:

$$\mathbf{R} = \frac{1}{\hat{\sigma}_0^2} \mathbf{P} \mathbf{C}_{\hat{\mathbf{v}}\hat{\mathbf{v}}} \quad 8.58$$

chiamando:

$$\hat{p}_i = \frac{\sigma_{v_i}^2}{\hat{\sigma}_0^2}$$

i pesi degli scarti dopo la compensazione.

Se \mathbf{P} è diagonale si ha:

$$r_i^i = \frac{\hat{p}_i}{\hat{p}_i} \quad 8.59$$

Attraverso la 8.58 e l'espressione della $\mathbf{C}_{\hat{\mathbf{v}}\hat{\mathbf{v}}}$ si ha:

$$\mathbf{R} = \mathbf{I} - \mathbf{P} \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T \quad 8.60$$

8.5 IL PRINCIPIO DEI MINIMI QUADRATI IN CASI NON LINEARI

Premettiamo che in casi non lineari il metodo perde le proprietà di ottimalità descritte in precedenza e può anche ammettere più soluzioni.

Siano date ℓ equazioni, funzioni delle osservabili y e dei parametri x :

$$\mathbf{g}(x, y) = \begin{Bmatrix} g_1(x, y) \\ g_2(x, y) \\ \vdots \\ g_\ell(x, y) \end{Bmatrix} = 0 \quad 8.61$$

con $y \in \mathbf{R}^m$ ed $x \in \mathbf{R}^n$

Si cercano le stime \hat{x} , \hat{y} tali che $(y_0 - \hat{y})^T P(y_0 - \hat{y}) = \min$, sotto la condizione $g(\hat{x}, \hat{y}) = 0$.

Supponiamo di conoscere i valori approssimati \tilde{x} , \tilde{y} e che, nell'intorno di detti valori, g sia ben linearizzabile, dimodoché:

$$\left. \begin{aligned} \hat{y} &= \tilde{y} + \eta \\ \hat{x} &= \tilde{x} + \xi \end{aligned} \right\} \quad 8.62$$

Linearizzando $g(\hat{x}, \hat{y}) = 0$ attorno ai valori approssimati si (\tilde{x}, \tilde{y}) ottiene:

$$g(\hat{x}, \hat{y}) = 0 \cong g(\tilde{x}, \tilde{y}) + \left(\frac{\partial \tilde{g}}{\partial x}\right)\xi + \left(\frac{\partial \tilde{g}}{\partial y}\right)\eta \quad 8.63$$

Chiamiamo con:

$$\frac{\partial g}{\partial x} = \left[\frac{\partial g_i}{\partial x_j} \right] = A \quad 8.64$$

la matrice disegno calcolata nei valori (\tilde{x}, \tilde{y}) ; con:

$$\frac{\partial g}{\partial y} = \left[\frac{\partial g_i}{\partial y_k} \right] = -D \quad 8.65$$

ed infine con:

$$g(\tilde{x}, \tilde{y}) = d \quad 8.66$$

Si arriva perciò al sistema linearizzato:

$$D\eta = A\xi + d \quad 8.67$$

notiamo che essendo:

$$\hat{y} = \tilde{y} + \eta$$

si ha:

$$C_{\eta\eta} = C_{\hat{y}\hat{y}} = \sigma_0^2 Q \quad 8.68$$

Si noti poi che:

$$(y_0 - \hat{y})^T P(y_0 - \hat{y}) = (y_0 - \tilde{y} - \tilde{\eta})P(y_0 - \hat{y} - \hat{\eta})$$

e, posto:

$$\eta_0 = y_0 - \tilde{y}$$

si ha da soddisfare:

$$(\eta_0 - \hat{\eta})^T P(\eta_0 - \hat{\eta}) = \min \quad 8.69$$

sotto la condizione 8.67.

Da questo punto in poi la soluzione è quindi analoga al caso lineare già visto. Dopo aver ricavato la soluzione in $\hat{\xi}$ ed in $\hat{\eta}$ si calcola il vettore degli scarti:

$$\hat{\mathbf{v}} = (\eta_0 - \hat{\eta}) \quad 8.70$$

Gli errori di chiusura valgono:

$$U = Dy_0 - A\hat{x} - d = D\eta_0 - A\hat{\xi} - d$$

Ricordando la 8.63 che si scrive anche:

$$g(\hat{x}, \hat{y}) = d + A\hat{\xi} - D\eta = 0$$

si ha:

$$U = g(\hat{x}, \hat{y}) = g(\tilde{x} + \hat{\xi}, \tilde{y} + \hat{\eta}) \quad 8.71$$

Se gli scarti sono elevati si itera il procedimento, a partire dalle stime \hat{x} ed \hat{y} , utilizzate ora come valori approssimati. Si prosegue nelle iterazioni sinché:

$$\sigma_{\eta(i+1)}^2 = \hat{\mathbf{v}}_{i+1}^T P \hat{\mathbf{v}}_{i+1} < \sigma_{\eta(i)}^2 \quad 8.72$$

Una seconda alternativa nella scelta di fermare o proseguire le iterazioni consiste nel verificare che le correzioni alle misure ed ai parametri sono trascurabili; scelto così un valore ε piccolo a piacere:

$$\|\hat{\xi}_{i+1} - \hat{\xi}_i\| < \varepsilon_1 \quad \text{oppure} \quad \|\hat{\eta}_{i+1} - \hat{\eta}_i\| < \varepsilon_2 \quad 8.73$$

Infine si osservi che se le funzioni $g(x, y)$ sono date in forma esplicita rispetto alle osservabili, cioè se sono del tipo:

$$y = g(x) \quad 8.74$$

non occorre linearizzare rispetto ad y ; questo è in realtà il caso nel quale riusciamo quasi sempre a ricondurre le equazioni nelle osservabili (equazioni generatrici).

Anche complicando un poco la funzione g , è preferibile ricondurci a questo approccio, perché più semplice da programmare: ci si riduce infatti al caso di osservazioni non lineari con soli parametri aggiuntivi.

Si noti ancora che, nel caso di equazioni lineari non occorre la conoscenza di parametri approssimati, cosa invece indispensabile in caso contrario.

Nella ricerca della trasformazione 8.74 in forma esplicita, se è possibile, occorre privilegiare la linearità della funzione a motivo delle proprietà di ottimalità descritte.

8.6 ESERCIZIO

Si desidera esaminare e risolvere il problema della rototraslazione con doppia variazione di scala (relativa cioè ai due assi) di un sistema ortogonale su un sistema non ortogonale (fig. 8.3).

Trascuriamo per semplicità espositive per ora l'effetto dovuto alla traslazione. Esaminiamo prima il modello geometrico, poniamo poi alcune semplici ipotesi su quello stocastico e risolviamo infine il problema ai minimi quadrati secondo la tecnica dei parametri aggiuntivi.

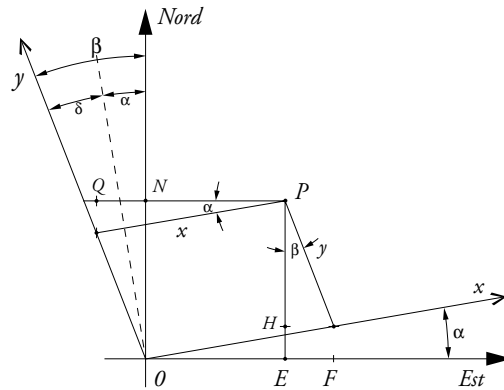


Fig. 8.3 – Trasformazione affine tra due sistemi di coordinate.

Modello geometrico

Consideriamo il punto P (fig. 8.3) di coordinate (E, N) nel sistema cartesiano ortogonale e di coordinate (x, y) nel sistema di assi non ortogonali.

Sia α l'angolo orario da x verso Est e β l'angolo da y verso Nord; chiamiamo «affinità» l'angolo $\delta = \beta - \alpha$.

Si avrà:

$$N = PH + HE = y \cos\beta + x \sin\alpha$$

$$E = OF - EF = x \cos\alpha - y \sin\beta$$

chiamando:

$$\cos\alpha = a, \quad \sin\alpha = c$$

$$-\sin\beta = b, \quad \cos\beta = d$$

si può scrivere il sistema lineare; si arriva alla stessa conclusione considerando l'espressione dei versori degli assi (x e y):

$$E = ax + by \quad \text{con le condizioni} \quad a^2 + c^2 = 1$$

$$N = cx + dy \quad \quad \quad \quad \quad \quad \quad b^2 + d^2 = 1$$

Ora avviene che, se si ipotizzano due fattori di scala per ciascun asse:

$$x = \lambda_x X$$

$$y = \lambda_y Y$$

$$E = a\lambda_x x + b\lambda_y y = (def) = AX + BY \tag{8.75}$$

$$N = c\lambda_x x + d\lambda_y y = (def) = CX + DY \tag{8.76}$$

mentre le condizioni di normalizzazione riportate sopra per A, B, C, D , divengono:

$$A^2 + C^2 = \cos^2\alpha \lambda_x^2 + \sin^2\alpha \lambda_x^2 = \lambda_x^2$$

$$B^2 + D^2 = \sin^2\beta \lambda_y^2 + \cos^2\beta \lambda_y^2 = \lambda_y^2$$

È possibile verificare inoltre che:

$$\frac{C}{A} = \frac{c}{a} = \operatorname{tg}\alpha$$

$$-\frac{B}{D} = -\frac{b}{d} = \operatorname{tg}\beta$$

Sfruttando queste relazioni e tenendo conto che:

$$\operatorname{tg}\delta = \operatorname{tg}(\beta - \alpha) = \frac{\sin\beta \cos\alpha - \cos\beta \sin\alpha}{\cos\beta \cos\alpha + \sin\beta \sin\alpha}$$

si ricava:

$$\operatorname{tg}\delta = \frac{AB + CD}{BC - DA}$$

Abbiamo sinora esaminato il modello fisico-geometrico senza traslazioni di assi, che ci ha portati nelle condizioni di risolvere un sistema di equazioni lineari 8.75 ed 8.76. Nel caso generale tuttavia rimane ancora da considerare una traslazione fra due sistemi; la 8.75 e la 8.76 divengono allora:

$$E = AX + BY + \Delta E \quad 8.77$$

$$N = CX + DY + \Delta N \quad 8.78$$

Se, di un numero n di punti dei quali sono note le coordinate (X, Y) , si sono misurate anche le coordinate (E, N) , le due equazioni possono essere scritte sinteticamente in forma matriciale:

$$\begin{pmatrix} E_i \\ N_i \end{pmatrix} = \begin{pmatrix} X_i & Y_i & 0 & 0 & 1 & 0 \\ 0 & 0 & X_i & Y_i & 0 & 1 \end{pmatrix} (A, B, C, D, \Delta E, \Delta N)^T \quad 8.79$$

Modello stocastico e soluzione ai minimi quadrati

È facile riconoscere qui sopra le misure y nelle coordinate E_i, N_i che ipotizziamo incorrelate, tali che $C_{EN} = \sigma^2 I$.

Riconosciamo poi i sei parametri incogniti nel vettore trasposto ed i coefficienti di questi parametri come matrice A .

Perché il problema possa avere una soluzione occorrerà avere a disposizione almeno tre coppie di coordinate in entrambi i sistemi. Applicando la 8.23 a questo esempio si nota che $a=0$.

Ipotizzando di scrivere dapprima tutte le equazioni nelle E_i e poi tutte quelle N_i

possono costruirsi la matrice disegno ed il vettore dei termini noti che assumono la forma:

$$l = y_0 = (E_i, N_i)^T$$

$$\underline{A} = \begin{Bmatrix} X_i & Y_i & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & Y_n & 0 & 0 & 1 & 0 \\ 0 & 0 & X_i & Y_i & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & X_n & Y_n & 0 & 1 \end{Bmatrix}; \quad l = \begin{Bmatrix} E_i \\ \vdots \\ E_n \\ N_i \\ \vdots \\ N_n \end{Bmatrix} \quad 8.80$$

Facciamo poi l'ipotesi semplificativa che:

$$\sigma_0^2 = \sigma_i^2 = 1; \quad P = I$$

La normalizzazione della 8.80

La costruzione della matrice normale 8.33 $N = A^T P A$, porta ad ottenere una matrice di dimensioni $N_6^6 = A^T A$ ed un vettore b di dimensione $6 \cdot b = A^T P l$.

$$N = \begin{pmatrix} \sum X_i^2 & \sum X_i Y_i & 0 & 0 & \sum X_i & 0 \\ \text{simm} & \sum Y_i^2 & 0 & 0 & \sum Y_i & 0 \\ & & \sum X_i^2 & \sum X_i Y_i & \sum X_i & 0 \\ & & & \sum Y_i^2 & \sum Y_i & 0 \\ & & & & n & 0 \\ & & & & & n \end{pmatrix}; \quad b = \begin{pmatrix} \sum X_i E_i \\ \sum Y_i E_i \\ \sum X_i N_i \\ \sum Y_i N_i \\ \sum E_i \\ \sum N_i \end{pmatrix}$$

È facile ottenere questo risultato per la matrice normale moltiplicando fra loro le colonne di A e moltiplicando le colonne di A e di l per i termini noti normalizzati.

Occorre ora risolvere un sistema lineare di sei equazioni in sei incognite, ma è possibile una ulteriore semplificazione. Si nota che, se con un artificio, rendessimo nulli alcuni termini: $\sum E_i = \sum N_i = \sum X_i = \sum Y_i = 0$ si semplificherebbe di molto il problema: scomparirebbero così le ultime righe e colonne di N . Ciò è possibile, se le coordinate nei due sistemi di partenza, che ora chiamiamo (X', Y') ed (E', N') sono tali che, calcolate le coordinate dei baricentri:

$$X_G = \frac{\sum X'}{n}; \quad Y_G = \frac{\sum Y'}{n}$$

$$E_G = \frac{\sum E'}{n}; \quad N_G = \frac{\sum N'}{n}$$

Si definiscano le coordinate X, Y, E, N in questo modo:

$$\begin{aligned} X &= X' - X_G; & Y &= Y' - Y_G \\ N &= N' - N_G; & E &= E' - E_G \end{aligned}$$

Così si avrà sempre che:

$$\sum X_i = \sum X' - nX_G = 0$$

e similmente:

$$\sum Y_i = \sum E_i = \sum N_i = 0$$

In questo modo il problema si riduce al calcolo di soli 4 parametri a due a due incorrelati: (A, B) e (C, D) :

$$\begin{pmatrix} \sum X_i^2 & \sum X_i Y_i & 0 & 0 \\ \text{Simm} & \sum Y_i^2 & 0 & 0 \\ & & \sum X_i^2 & \sum X_i Y_i \\ & & & \sum Y_i^2 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} \sum X_i E_i \\ \sum Y_i E_i \\ \sum X_i N_i \\ \sum Y_i N_i \end{pmatrix} \quad 8.81$$

che può essere diviso nei due sistemi:

$$\begin{bmatrix} N \\ N \end{bmatrix} \begin{pmatrix} A, B \\ C, D \end{pmatrix}^T = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad 8.82$$

Si ottiene facilmente la stima di questi parametri: chiamando con Δ il determinante della matrice N:

$$\Delta = \sum X_i^2 \sum Y_i^2 - (\sum X_i Y_i)^2$$

Le formule:

$$\begin{aligned} \hat{A} &= (\sum Y_i^2 \sum X_i E_i - \sum X_i E_i \sum Y_i E_i) / \Delta \\ \hat{B} &= (-\sum X_i Y_i \sum X_i E_i + \sum X_i^2 \sum Y_i E_i) / \Delta \\ \hat{C} &= (\sum Y_i^2 \sum X_i N_i - \sum X_i Y_i \sum Y_i N_i) / \Delta \\ \hat{D} &= (-\sum X_i Y_i \sum X_i N_i + \sum X_i^2 \sum Y_i N_i) / \Delta \end{aligned}$$

risolvono il problema. Si ricava poi, ricordando le 8.5, 8.25 e 8.26:

$$\begin{aligned} \hat{\mathbf{v}} &= y_0 - A\hat{x} \\ \hat{\mathbf{v}}_{Ei} &= E_i - x_i - \hat{A} - Y_i \hat{B} \\ \hat{\mathbf{v}}_{Ni} &= N_i - X_i \hat{C} - Y_i \hat{D} \end{aligned}$$

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T \mathbf{v}}{2n - 4}$$

Utilizzando poi la 8.45 si ha:

$$C_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 \mathbf{N}^{-1}$$

cioè:

$$\sigma_{\hat{A}}^2 = \hat{\sigma}_0^2 \sum Y_i^2 / \Delta = \sigma_{\hat{C}}^2$$

$$\sigma_{\hat{A}\hat{B}}^2 = -\hat{\sigma}_0^2 \sum X_i Y_i / \Delta = \sigma_{\hat{C}\hat{D}}^2$$

$$\sigma_{\hat{B}}^2 = \hat{\sigma}_0^2 \sum X_i^2 / \Delta = \sigma_{\hat{D}}^2$$

Ricavando poi:

$$C_{\hat{y}\hat{y}} = \hat{\sigma}_0^2 \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T$$

si nota ancora che le varianze delle coordinate \hat{E}, \hat{N} ricavabili da X e Y sono identiche e valgono:

$$\sigma_{\hat{E}}^2 = \sigma_{\hat{N}}^2 = \sigma_{\hat{A}}^2 X^2 + \sigma_{\hat{B}}^2 Y^2 + 2\sigma_{\hat{A}\hat{B}}^2 XY$$

mentre sono nulle le covarianze $\sigma_{\hat{E}\hat{N}}$ sempre ammessa l'ipotesi di partire da una matrice dei pesi proporzionale alla matrice identità.

