Indice

1 Introduzione alla problematica dei fenomeni non deterministici

Si considerino i seguenti esperimenti semplici:

a) una moneta con le facce contrassegnate da testa e croce viene lanciata a mano su un tavolo e si osserva la faccia che resta scoperta una volta che la moneta si è fermata;

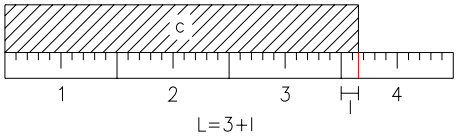


Figura 1.1

b) dato un corpo rigido C di lunghezza L di oltre 3 metri e dato un metro campione suddiviso in decimetri, centimetri e millimetri, si misura L con l'approssimazione del millimetro mediante il metodo del riporto (fig. 1.1);

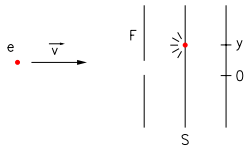


Figura 1.2

c) sia dato un elettrone con velocità $\vec{v}$ che viene fatto passare attraverso una fessura F (fig. 1.2); con uno schermo fluorescente S si riveli la posizione in cui l'elettrone viene a cadere, misurando in particolare la coordinata y.

Questi tre esperimenti hanno un aspetto fondamentale in comune: in base ai dati a nostra disposizione è impossibile predire con esattezza il risultato dell'esperimento.

Un altro modo per esprimere lo stesso concetto è: ripetendo più volte lo stesso esperimento, pur mettendosi il più possibile nelle "stesse" condizioni, si ottengono risultati diversi.

Nel caso a) ciò è dovuto al fatto che una piccolissima variazione nella spinta iniziale, nella velocità di rotazione e nella posizione relativa manotavolo comporta, dopo la caduta, una posizione completamente diversa della moneta. Naturalmente se il lancio avvenisse in condizioni *perfettamente uguali* (ad esempio con un dispositivo meccanico di lancio e sotto vuoto) il risultato sarebbe ogni volta identico, ma se il lancio è fatto a mano è impossibile raggiungere un tale controllo dell'impulso e della posizione iniziale in modo da ottenere un risultato fisso.

Nel caso b) l'indeterminatezza del risultato è dovuta sia alla difficoltà di avere un perfetto allineamento tra l'estremo di C e l'origine del metro nella posizione 1, sia alla difficoltà di riportare l'origine del metro nelle posizioni 2, 3, 4 nei punti in cui si aveva la fine dello stesso nelle posizioni 1, 2, 3, sia infine alla difficoltà di arrotondare al millimetro la lettura ℓ eseguita nella posizione 4.

Dopo aver fatto qualche prova si troverà che, procedendo con attenzione, si ottengono misure di L che differiscono tra loro di 1 mm o, in casi eccezionali, di 2 mm.

Si noti che se avessimo preteso, con lo stesso metodo, di stimare L al decimo di millimetro, avremmo ottenuto dei numeri assai più variabili (letture differenti per parecchi decimi di millimetro), mentre se avessimo richiesto di stimare L al centimetro il risultato sarebbe stato sempre uguale; da questa semplice osservazione ricaviamo l'indicazione che il problema della indeterminazione si presenta nella misura in cui si cerchi di spingere l'approssimazione ai limiti delle capacità dell'apparato di misura stessa.[1]

Nel caso c) l'indeterminazione di y è dovuta ad un principio generale,

[1] Nell'esempio in esame gioca un ruolo importante anche il grado di complessità dell'esperimento (nel nostro caso il numero di riporti); ad una maggiore complessità infatti corrisponde una maggiore indeterminazione del risultato. Si pensi infatti che se L fosse stata inferiore al metro, sarabbe stato possibile ottenere misure sempre uguali nell'ambito dell'arrotondamento al millimetro.

accettato come legge della natura, secondo il quale nel momento in cui la coordinata y è obbligata ad essere compresa in un intervallo (ciò avviene quando l'elettrone passa attraverso F), in corrispondenza la componente v_y della velocità (che prima era zero), subisce una indeterminazione tanto più ampia quanto più stretta era la fessura: si tratta del celebre principio di indeterminazione di Heisenberg.

Come si vede in tutti e tre i casi, anche se per motivi diversi, è impossibile descrivere ciò che ci si aspetta come risultato degli esperimenti, per mezzo di semplici variabili numeriche, deterministiche. In tutti e tre i casi alla domanda "quale risultato ti aspetti dall'esperimento?" sulla base di esperienze precedenti, si può solo rispondere ipotizzando una gamma di possibili valori e un ordine di priorità tra di essi. Questa priorità espressa mediante un numero reale compreso tra 0 e 1 prende il nome di probabilità.

2 La definizione di probabilità

Riassumendo la discussione del §1 possiamo dire che vi sono esperimenti il cui risultato non è univocamente definito: si ha piuttosto un insieme di possibili risultati e un ordine di priorità tra di essi.

Occorre creare degli enti matematici che rappresentino in modo formale tali proprietà. Questi enti sono le variabili causali che possiamo costruire nel seguente modo:

- indichiamo con S l'insieme dei possibili risultati dell'esperimento in oggetto; per semplicità conviene rappresentare S mediante un insieme di punti (in uno spazio a 1 o più dimensioni);
- consideriamo una parte di tali risultati, ossia un sottoinsieme di S, caratterizzati dal fatto di godere tutti di una proprietà q;
- stabiliamo una regola che a tutti i sottoinsiemi, ovvero a tutte le proprietà q, che ci interessano, assegna un ordine di priorità espresso mediante un numero reale compreso tra 0 e 1.

Questa regola prende il nome di distribuzione di probabilità. Naturalmente a seconda di come tale regola viene stabilita, in rapporto alle proprietà fisiche dell'esperimento, si ottengono diversi schemi concettuali di rappresentazione della realtà sperimentale.

Riportiamo qui tre diverse definizioni di probabilità.

I La definizione di Laplace

Questo autore si rifà ad un concetto di simmetria, suddividendo l'insieme dei possibili risultati di un esperimento in gruppi tra loro indistinguibili a priori per motivi di simmetria; ad ogni gruppo viene poi assegnata la stessa probabilità.

Ad esempio, per l'esperimento lancio di una moneta, se è impossibile distinguere fisicamente una faccia dall'altra, si porrà

$$\begin{aligned} &\text{probabilità di avere testa } = 1/2 \\ &\text{probabilità di avere croce } = 1/2 \end{aligned}$$

Se l'esperimento fosse "due lanci consecutivi della moneta" si avrebbero i seguenti possibili risultati:

$$\text{cc, ct, tc, tt}$$

Essendo indistinguibili a priori, ad ognuno di essi si assegnerà probabilità $= 1/4$; se poi si volesse definire la probabilità che su due lanci si ottenga croce almeno una volta, si può osservare che tre dei risultati possibili soddisfano tale requisito e si assume perciò:

$$\text{probabilità di ottenere una croce su due lanci } = 3/4$$

La difficoltà di tale definizione sta nel fatto che non sempre si può decomporre l'insieme dei possibili risultati di un esperimento non deterministico in base ad un principio di simmetria.

II La definizione di Von Mises

Questa definizione prende lo spunto da un principio, stabilito su basi sperimentali, che va sotto il nome di *legge empirica del caso*.

Si considerino N ripetizioni dello stesso esperimento e si conti il numero N_q dei risultati che verificano una certa proprietà q: ad esempio, nel caso

b), si prendano tutte le misure di L [2] uguali ad un valore assegnato. Il rapporto N_q/N prende il nome di frequenza relativa della proprietà q, sulle N prove: si verifichi sperimentalmente che all'aumentare di N, tale frequenza, pur variando da un N ad un altro, tende ad assumere un valore nel senso che le oscillazioni più ampie attorno ad esso tendono a diventare sempre più rade. Un esempio tipico è proprio quello del lancio di una moneta "non truccata": in fig. 2.1 riportiamo la sequenza di frequenze di "croci" su 50 lanci.

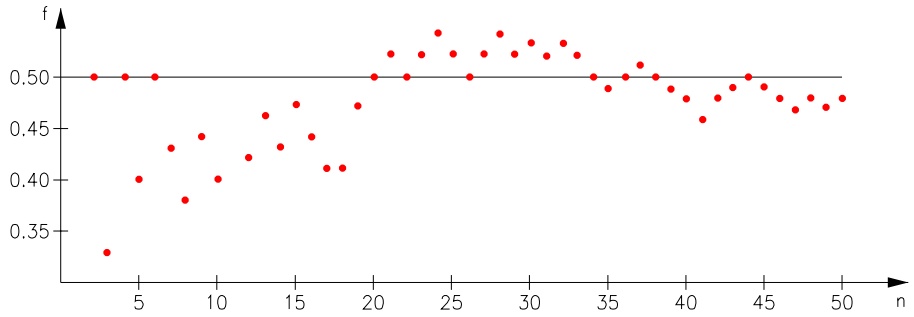


Figura 2.1

Assumendo per principio che il limite esista, si pone allora

$$\text{(probabilità che il risultato dell'esperimento verifichi la proprietà } q) = \lim_{N\to\infty} \frac{N_q}{N}$$

Benché molto attraente, perché stabilisce un legame diretto tra esperienza e schema teorico, questa definizione presenta una difficoltà di ordine logico: infatti si può notare che, benché estremamente improbabile dal punto di vista intuitivo, non è impossibile che ogni lancio di una lunga serie dia croce come risultato.

Sulla base di questa serie di esperienze si arriverebbe perciò a concludere che

[2] Arrotondate al millimetro.

$$\text{probabilità di ottenere croce } = 1.$$

Naturalmente bisognerebbe essere molto "sfortunati" per incappare in un tale caso, ma ciò non è impossibile a priori.

III La definizione assiomatica di probabilità

Si tratta della definizione oggi usata e sarà da questo punto di vista che ci porremo nello studio delle variabili casuali nei prossimi paragrafi.

Questo punto di vista consiste nel definire le distribuzioni di probabilità in base alle proprietà assiomatiche cui esse devono soddisfare; per non dilungarci eccessivamente diciamo che una distribuzione di probabilità sull'insieme S, che prende il nome di insieme dei valori argomentali, non é altro che una misura, su una famiglia di sottoinsiemi, $\mathcal{A}$, di S che include S stesso e l'insieme vuoto $\emptyset$, che oltre ai normali assiomi della misura soddisfa anche la relazione

$$P(S) = 1 \quad . \tag{2.1}$$

Notiamo che la misura P è tale rispetto alla famiglia $\mathcal{A}$ che ha il ruolo della famiglia degli insiemi misurabili secondo P. Ricordiamo che per essere una misura P deve soddisfare le relazioni

$$P(A) \geq 0 \ , \ P(\emptyset) = 0 \tag{2.2}$$

$$P(A \cup B) = P(A) + P(B) \quad (A \cap B = \emptyset) \tag{2.3}$$

$$P(A^C) = 1 - P(A) \quad (A^C = S - A) \tag{2.4}$$

Già da questa prima proprietà si comprende come deve essere strutturata la famiglia degli insiemi misurabili. Ad esempio supporremo per ipotesi di poter sempre unire due insiemi (o eventi) A e $B \in \mathcal{A}$ ed ottenere ancora un insieme di $\mathcal{A}$ (cioè misurabile), anche quando A e B non siano disgiunti come in (2.3). A questo punto appare ovvio anche che l'intersezione di due eventi A e B deve essere misurabile, infatti

$$A \cap B = (A^C \cup B^C)^C \quad . \tag{2.5}$$

Pertanto la famiglia $\mathcal{A}$ deve almeno essere un'algebra di insiemi, munita di unità, cioè S per cui vale $A \cap S = A,\ \forall A \in \mathcal{A}$.

Questo insieme di proprietà, a rigore, definisce uno schema probabilistico in senso lato; infatti per motivi tecnici è spesso necessario in teoria della probabilità fare non solo l'unione di 2 (o comunque di un numero finito) eventi, ma anche l'unione di una successione di eventi A_n. richiedendo che anche $\bigcup_{n=1}^{+\infty} A_n \in \mathcal{A}$.

La proprietà corrispondente in termini di probabilità deve essere che

$$P(\bigcup_{n=1}^{+\infty} A_n) = \sum_{n=1}^{+\infty} P(A_n) \tag{2.6}$$

quando gli A_n sono a due a due disgiunti, cioè

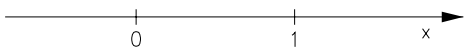
$$A_n \cap A_m = \emptyset \qquad n \neq m \quad . \tag{2.7}$$

Queste specificazioni servono solo se l'insieme S è infinito, mentre quando S è finito uno schema in senso lato è tutto ciò di cui si ha bisogno.

Ad esempio nel caso del lancio di una moneta, S è costituito da due valori argomentali che per convenzione possiamo porre

$$\begin{aligned} X = 0 \quad & \text{corrisponde al risultato testa} \\ X = 1 \quad & \text{corrisponde al risultato croce} \quad ; \end{aligned}$$

S è dunque l'insieme dei due punti di coordinate 0 e 1.



$$(S = \{0\} \quad \{1\}) \quad .$$

I soli sottoinsiemi di S sono

$$\emptyset, \quad \{0\}, \quad \{1\}, \quad \{0,1\} \ ;$$

se la moneta non è truccata, seguendo anche l'indicazione della definizione di Laplace, possiamo porre

$$\begin{array}{lcl} P(\emptyset) & = & 0 \\ P(\{0\}) & = & 1/2 \\ P(\{1\}) & = & 1/2 \\ P(\{0,1\}) & = & 1 \end{array}$$

In questo modo gli assiomi (2.1), (2.2), (2.3) sono soddisfatti, e diremo che è definita una distribuzione di probabilità.

E' facile generalizzare questo esempio al caso generale di un insieme finito, costituito da N punti per il quale si dimostra che il numero di possibili sottoinsiemi è 2^N.

Naturalmente il vero problema è: che rapporto esiste tra questa definizione e il fatto concreto, ad esempio che se lancio N volte una moneta ottengo N_t volte testa ed N_c volte croce?

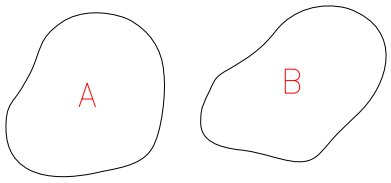
Grande merito di questo approccio assiomatico alla teoria della probabilità é proprio quello di essere in grado, a partire dai semplici assiomi detti, di definire degli enti teorici il cui comportamento è marcatamente simile a quello trovato empiricamente per N_t ed N_c; in particolare, come si vedrà nei prossimi paragrafi, il rapporto $\frac{N_c}{N}$ = (frequenza di croci su N lanci), tenderà ad avere per N crescente un andamento che ben spiega il comportamento visto in fig. 2.1. In questo senso la giustificazione dello schema matematico creato si ha a posteriori nel confronto tra risultati teorici e risultati empirici.

3 Primi teoremi sulle distribuzioni di probabilità

a) Teorema della probabilità totale

Si vuole risolvere il seguente problema: dati due eventi A e B (due sottoinsiemi di S) quale è la probabilità che si verifichi A o B, cioè quanto vale $P(A \cup B)$?

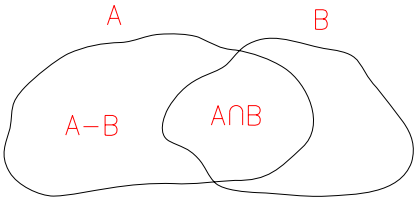
Se i due insiemi A e B sono disgiunti, per la definizione assiomatica di probabilità



$$\begin{aligned} (A \cap B &= \emptyset) \qquad (3.1) \\ P(A \cup B) &= P(A) + P(B) \quad . \end{aligned}$$

Figura 3.1

Se invece A e B non sono disgiunti si può notare che



$$A \cup B = (A - B) \cup B$$

Figura 3.2

e poiché evidentemente $A - B$ e B sono disgiunti

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

Ora notiamo che

$$A = (A - B) \cup (A \cap B)$$

e poiché ancora $A - B$ e $A \cap B$ sono disgiunti, si ha

$$P(A) = P(A - B) + P(A \cap B) \quad .$$

Ricavando da questa $P(A - B)$ e sostituendo nella precedente si trova il teorema della probabilità totale:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (3.2)$$

Si osservi che la (3.2) contiene anche la (3.1).

Esempio 3.1: estraendo a caso una carta da un mazzo da scopa (40 carte) si vuole sapere quale è la probabilità che la carta sia cuori oppure che sia una figura. In questo caso gli eventi A e B sono

$$\begin{aligned} &\text{la carta estratta è cuori } = A \\ &\text{la carta estratta è una figura } = B \end{aligned}$$

A è formato dall'unione di 10 eventi distinti (estraggo l'asso, il due, ... il re di cuori), ciascuno dei quali ha probabilità $P = 1/40$; perciò

$$P(A) = 10/40 = 0,250 \quad .$$

B invece è formato da 12 eventi distinti ed equiprobabili (fante, donna, re di cuori; fante, donna, re di fiori, ecc.) perciò

$$P(B) = 12/40 = 0,300$$

$A \cap B$ è formato da 3 eventi distinti ed equiprobabili (fante, donna, re di cuori) e quindi

$$P(A \cap B) = 3/40 = 0,075 \quad .$$

La risposta è quindi

$$P(A \cup B) = 0,250 + 0,300 - 0,075 = 0,475$$

b) Definizione di probabilità condizionata

Spesso accade che sia importante esaminare la distribuzione di probabilità di una certa variabile su una parte soltanto dei valori argomentali, cioè restringendo l'insieme S ad un suo sottoinsieme. Ad esempio si consideri una popolazione di 100 individui classificati secondo il colore degli occhi e dei capelli:

		capelli	
		chiari	scuri
occhi	chiari	40	10
	scuri	10	40

se estraggo un individuo a caso, la probabilità che questo abbia occhi chiari sarà data da

$$P = (40 + 10)/100 = 0,5.$$

Ora però posso chiedermi, se separo gli individui dai capelli chiari da quelli coi capelli scuri, cosa vale la probabilità di estrarre a caso un individuo con occhi chiari, tra quelli con i capelli chiari?

La risposta è evidentemente

$$P = 40/50 = 0,8$$

Come si vede, isolando una parte della popolazione con dati valori argomentali (nell'esempio restringendoci ai soli individui con capelli chiari) si genera una nuova distribuzione di probabilità.

Indirizzati dall'esempio, poniamo per definizione come probabilità dell'evento A, condizionato dall'evento B,

$$P(A|B) = P(AB)/P(B) \quad [3] \tag{3.3}$$

$P(A|B)$ risponde alla domanda: quale è la probabilità che il risultato del nostro esperimento, ovvero una estrazione della nostra variabile, appartenga ad A, dando per scontato che questo verifica la proprietà B (cioè appartiene a B)?

E' facile dimostrare formalmente che $P(A|B)$ è una vera distribuzione di probabilità con insieme dei valori argomentali uguale a B.

Nell'esempio usato precedentemente se si indicano con:

$A =$ estrazione di un individuo con occhi chiari
$B =$ estrazione di un individuo con capelli chiari

si ha proprio

$$\begin{array}{lll} P(B) & = 50/100 & = 0,5 \\ P(AB) & = 40/100 & = 0,4 \\ P(A|B) & = 0,4/0,5 & = 0,8 \end{array}$$

[3] Per semplicità indichiamo l'intersezione tra i due sistemi A e B con la formula $A \cap B = AB$.

come si era già trovato intuitivamente (aiutandoci con la definizione di Laplace).

La prima cosa che salta all'occhio proprio in questo esempio è che $P(A|B) = 0,8$ è molto più alta che $P(A) = 0,5$; questo significa che esiste un qualche legame, nella popolazione in esame, tra il fatto di avere i capelli chiari e quello di avere occhi chiari. Questo legame non è deterministico, infatti non tutti gli elementi con capelli chiari hanno occhi chiari, ma probabilistico.

Questa considerazione induce a introdurre la *definizione di indipendenza stocastica di un evento A da un altro B:*

diciamo che A è stocasticamente indipendente da B se

$$P(A|B) = P(A) \qquad (3.4)$$

Esempio 3.2: nel caso visto nell'Esempio 3.1, A è stocasticamente dipendente da B, infatti

$$P(A|B) = 0,8 \neq P(A) = 0,5$$

Esempio 3.3: si consideri una popolazione di 100 individui catalogati per sesso e colore dei capelli.

		capelli	
		chiari	scuri
sesso	maschi	20	20
	femmine	30	30

Siano:

$A =$ estrazione di un individuo con capelli chiari
$B =$ estrazione di un individuo di sesso femminile

Risulta:

$$\begin{aligned} P(A) &= 50/100 = 0,5 \\ P(A|B) &= \frac{P(AB)}{P(B)} = \frac{30/100}{60/100} = \frac{0,3}{0,6} = 0,5 \end{aligned}$$

perciò A è stocasticamente indipendente da B.

c) Teorema della probabilità composta

Viene indicata come probabilità composta di A e B la probabilità che una estrazione appartenga contemporaneamente ad A e a B, cioè $P(AB)$.

Vale il seguente teorema:

> condizione necessaria e sufficiente perché due eventi A e B siano stocasticamene indipendenti è che la probabilità composta di A e B sia uguale al prodotto delle singole probabilità di A e B

$$P(AB) = P(A)P(B) \quad . \tag{3.5}$$

Infatti se A è indipendente da B

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

da cui segue (3.5); se viceversa (3.5) è verificata, allora

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

ed A risulta indipendente da B.

Osservazione 3.1: se A è indipendente da B allora B è indipendente da A; infatti, per la (3.5),

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

Osservazione 3.2: dalla definizione di probabilità condizionata si ricava la formula, sempre valida,

$$P(AB) = P(A|B)P(B) = P(B|A)P(A)$$

Da questa a sua volta si può ricavare

$$P(ABC) = P(A|BC)P(BC) = P(A|BC)P(B|C)P(C) \tag{3.6}$$

e così via.

d) Il teorema di Bayes

Questo teorema dal punto di vista formale può apparire come un semplice gioco di formule, in realtà è un primo risultato profondo su cui si basa un'intera branca della statistica, detta appunto Bayesiana.

L'idea di fondo è la seguente: sia $\{D_i, 1 = 1, 2 \dots n\}$ una partizione di S, ovvero

$$\begin{aligned} &\bigcup_{i=1}^{n} D_i = S \\ &D_i \cap D_j = \emptyset \quad i \neq j. \end{aligned} \tag{3.7}$$

Supponiamo che sia assegnata una distribuzione di probabilità su S per cui $P(D_i)$ sono note. Supponiamo ora di compiere un esperimento stocastico, descritto da P su S, di cui non si conosca esattamente il risultato, ma si sappia piuttosto che il punto che lo rappresenta deve stare in un qualche $A \subset S$, come nella figura seguente.

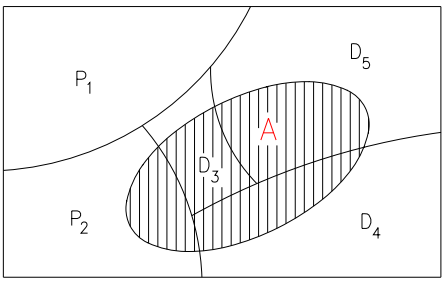


Figura 3.3

Ci si chiede, come ha influenzato la distribuzione di probabilità tra i D_i il fatto di sapere che l'esperimento ha un risultato vincolato ad A?

Teorema di Bayes: vale la formula

$$P(D_i|A) = \frac{P(A|D_i)P(D_i)}{\sum_{k=1}^{n} P(A|D_k)P(D_k)} \qquad (3.8)$$

In effetti la situazione sopra descritta corrisponde ad un restringimento della popolazione su cui si distribuiscono i risultati, così che le probabilità $P(D_i|A)$ sono la risposta al quesito. Che queste siano date dalla (3.8) deriva dall'osservare che

$$P(A|B)P(B) = P(B|A)P(A) = P(AB) \qquad (3.9)$$

così che

$$P(D_i|A) = \frac{P(A|D_i)P(D_i)}{P(A)} \quad . \qquad (3.10)$$

Ma, per la (3.7),

$$A = A \cdot (\bigcup_{k=1}^{n} D_{1k}) = \bigcup_{k=1}^{n} AD_k \quad ,$$

dove $AD_k \cap AD_n = 0, \ k \neq n$. Perciò

$$P(A) = \sum_{k=1}^{n} P(AD_k) = \sum_{k=1}^{n} P(A|D_k)P(D_k) \quad ,$$

che, sostituita nella (3.10), dà la (3.8).

Esempio 3.4: una rete di telecomunicazione è composta dagli elementi raffigurati qua sotto

G = generatore di un segnale
D = deviatore, può incanalare il segnale sul ramo 1 o sul ramo 2
I_1, I_2 = interruttori, che possono essere chiusi o aperti
R = rivelatore di segnale.

Il sistema è stocastico in quanto D può assumere casualmente i valori $D = 1, \ D = 2$ con probabilità

$$D = \begin{cases} 1 & 2 \\ \alpha_1 & \alpha_2 \end{cases} \quad (\alpha_1 + \alpha_2 = 1)$$

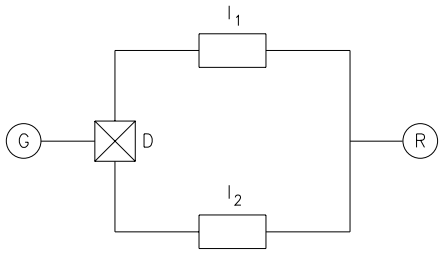


Figura 3.4

ed anche gli interruttori I_i possono essere chiusi o aperti a caso, con probabilità

$$I_i = \begin{cases} 0 \text{ (chiuso)} & 1 \text{ (aperto)} \\ p_i & q_i \end{cases} \quad (p_i + q_i = 1,\ i = 1,2)$$

Gli eventi $\{D = 1\}, \{D = 2\}, \{I_i = 0\}, \{I_i = 1\}$ sono per ipotesi tutti stocasticamente indipendenti tra loro. Perciò lo stato della rete è caratterizzato da 8 possibili valori $\{D = 1,2;\ I_1 = 0,1;\ I_2 = 0,1\}$ cui facciamo corrispondere le rispettive probabilità ricavate dal teorema delle probabilità composte

$$P\{D = 1, I_1 = k, I_2 = j\} = P\{D = 1\}P\{I_1 = k\}P\{I_2 = i\} \quad ,$$

I_2 \ I_1	0	1
0	p_1p_2	q_1p_2
1	p_1q_2	q_1q_2

$$X \begin{cases} \alpha_1\ ,\ D = 1 \\ \alpha_2\ ,\ D = 2 \end{cases}$$

L'esperimento stocastico in oggetto consiste nel connettere G ed osservare R; diciamo che R possa assumere i valori 0,1 a seconda che si riveli un segnale o no,

$$R = \begin{cases} 0 & 1 \\ P & Q \end{cases} \quad P + Q = 1 \quad .$$

Osserviamo che

$$\begin{aligned} P = P(R=0) &= P\{D=1, I_1=0, \forall I_2\} + P\{D=2, \forall I_1, I_2=0\} = \\ &= \alpha_1 p_1 + \alpha_2 p_2 \\ Q = P(R=1) &= 1 - P = \alpha_1 q_1 + \alpha_2 q_2 \quad . \end{aligned}$$

Ora supponiamo che l'esperimento sia eseguito ed effettivamente si veda che $R=1$, cioè si rivela un segnale. Ci si può chedere quale sia in questo caso la probabilità che $D=1, D=2$, cioè che il segnale sia passato per il ramo 1 o il ramo 2. Con il teorema di Bayes abbiamo

$$\begin{aligned} P(D=i|R=1) &= \frac{P(R=1|D=i)P(D=i)}{P(R=1)} = \\ &= \frac{\alpha_i q_i}{\alpha_1 q_1 + \alpha_2 q_2} \quad i=1,2 \quad . \end{aligned}$$

Notiamo che se tutto è simmetrico $\alpha_i = 1/2$, $p_i = q_i = 1/2$ risulta anche $P(D=1|R=1) = 1/2$, cioè non c'é alcun motivo di pensare che sia più probabile il passaggio per l'uno o l'altro dei rami. Se invece si ha qualche dissimmetria, l'informazione $R=1$ modifica la nostra conoscenza di D; ad esempio se $\alpha_1 = \alpha_2 = 1/2$ ma $q_1 = 3/4,\ q_2 = 1/2$ allora

$$P(D=1|R=1) = 3/5 \ , \ P(D=2|R=1) = 2/5$$

che è significativamente differente dalla conoscenza a priori di D.

Questo esempio, pur nella evidente schematizzazione, è rilevante per lo studio del traffico di informazioni in una rete di calcolatori.

4 Variabili casuali, variabili statistiche (a una dimensione)

a) Una variabile casuale a una dimensione è una variabile che descrive i possibili risultati di un esperimento stocastico per mezzo di una distribuzione di probabilità il cui insieme di valori argomentali S sia rappresentabile sulla retta reale e tale che sia definita la probabilità per ogni insieme del tipo

$$I(x_0) = \{x \leq x_0\} \cap S \quad . \tag{4.1}$$

La variabile casuale a una dimensione sarà perciò caratterizzata dalla funzione

$$F(x_0) = P[X \in I(x_0)] \ . \tag{4.2}$$

$F(x_0)$ prende il nome di *funzione di distribuzione* e gode delle seguenti proprietà:

1) $F(x_0)$ è definita per ogni x_0 reale e risulta $0 \leq F(x_0) \leq 1$

2) $\lim_{x_0 \to -\infty} F(x_0) = 0$

3) $\lim_{x_0 \to +\infty} F(x_0) = 1$

4) $F(x_2) \geq F(x_1), \quad \forall x_2 \geq x_1.$

Una variabile casuale si dice discreta se l'insieme S è formato da un insieme discreto di punti, su cui è concentrata la probabilità; se viceversa la probabilità che x assuma un singolo valore numerico è sempre zero, allora la variabile si dice continua.

Nel primo caso la funzione di distribuzione $F(x)$ è una funzione a gradini con discontinuità in corrispondenza ai valori argomentali, nel secondo caso $F(x)$ è continua.

Nel seguito useremo anche l'abbreviazione v.c. per indicare una variabile casuale.

Esempio 4.1: come si è già visto, il lancio di una moneta non truccata può essere rappresentato dalla variabile discreta:

$$\begin{cases} x_1 = 0 & x_2 = 1 & \text{valori argomentali} \\ p = 1/2 & q = 1/2 & \text{probabilità} \end{cases}$$

La funzione di distribuzione di tale variabile è

Esempio 4.2: si consideri una distribuzione di probabilità sull'intervallo $[0, 1]$, definito da

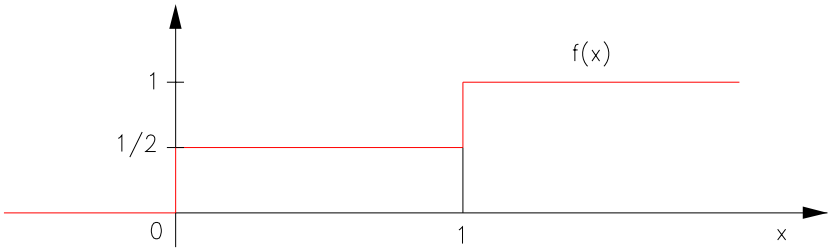


Figura 4.1

$$P(a \leq x \leq b) = b - a$$

(distribuzione uniforme).

La sua funzione di distribuzione sarà

$$\begin{aligned} F(x) &= 0 \quad , \quad x \leq 0 \\ F(x) &= x \quad , \quad 0 \leq x \leq 1 \\ F(x) &= 1 \quad , \quad x \geq 1 \end{aligned}$$

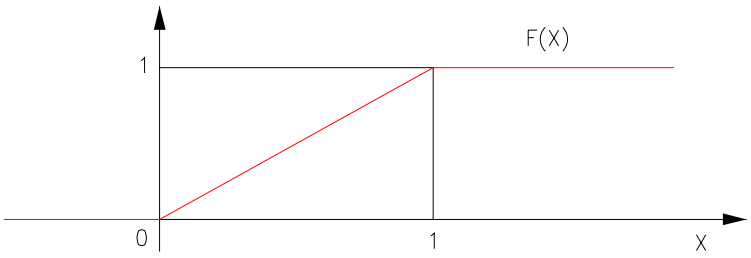


Figura 4.2

Come si vede, si tratta di una variabile continua.

Esempio 4.3: si definisca una distribuzione di probabilità tra $-\infty$ e $+\infty$, per mezzo della relazione

$$P(a \leq X \leq b) = \frac{1}{\pi}(\arctan b - \arctan a).$$

La corrispondente funzione di distribuzione risulta

$$F(x) = \frac{1}{\pi}\arctan x + \frac{1}{2}$$

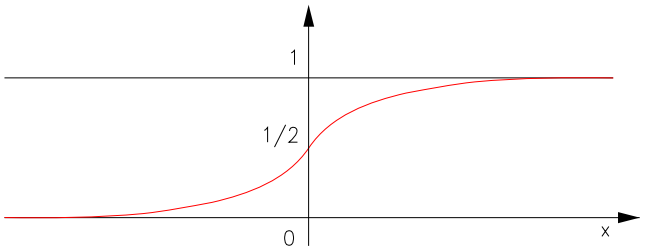


Figura 4.3

Osservazione 4.1: come si può vedere anche dagli esempi fatti, oltre che dalla definizione, la funzione di distribuzione può avere discontinuità solo nei punti in cui sia concentrata una probabilità finita: il salto della $F(x)$ corrisponde allora alla probabilità che X assuma quello specifico valore.

Osservazione 4.2: è bene comprendere che una v.c. data ha sempre una sua funzione di distribuzione, ma una stessa funzione di distribuzione può riferirsi a variabili casuali completamente diverse tra loro; così la stessa funzione a gradini dell'Esempio 4.1 può riferirsi a due diversi ed indipendenti giocatori a "testa e croce", il primo descritto da una v.c. X_1 ed il secondo da una v.c. X_2. Le estrazioni da X_1 ed X_2 risultano stocasticamente indipendenti, cioè legate a cause stocastiche che non interagiscono tra loro, nel senso del §3, sebbene ognuno dei due giocatori giudichi di avere la stessa funzione di distribuzione. All'estremo opposto abbiamo il caso in cui due v.c. apparentemente differenti, non solo hanno la stessa funzione di distribuzione ma risultano non distinguibili dal punto di vista probabilistico. Così si consideri ad esempio un dado con le facce

numerate da 0 a 5 ma munito di contrappesi tali per cui risultino sempre estratti lo 0 e l'1 ad ogni lancio, tranne che in casi assai particolari ed instabili da risultare descritti come eventi con probabilità nulla.

Dunque la funzione di distribuzione della v.c. corrispondente può essere descritta dicendo che si è in presenza di una variabile con valori argomentali a probabilità date da

$$X \equiv \begin{cases} 0 & 1 & 2 & 3 & 4 & 5 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \end{cases}$$

$$F(x) = 0, \quad 1/2 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1$$

La stessa funzione di distribuzione si avrebbe considerando una v.c. X', ridotta escludendo completamente i valori $X' = 2, 3, 4, 5$. In questo esempio dunque abbiamo due v.c. che apparentemente differiscono perché la prima, X, ha 6 valori argomentali, mentre la seconda, X', ne ha solo due. Tuttavia vale, proprio per costruzione, la fondamentale eguaglianza

$$P\{X = X'\} = 1 \ ,$$

così che le due v.c. sono dette *equivalenti*.

E' chiaro che dal punto di vista del calcolo della probabilità l'uso di X e di X' risulterà indifferente. Si comprende dunque come data una v.c. X a una dimensione, con funzione di distribuzione $F(x)$, si possa equivalentemente considerare al posto di X una sua estensione che ammetta come valori argomentali tutti quelli della retta reale, pur di aggiungere l'insieme $S^C = R - S$, e quindi tutti i suoi sottoinsiemi, a probabilità nulla.

b) Se per mezzo della variabile casuale si vuole rappresentare l'insieme dei possibili risultati di un esperimento non deterministico, altrettanto importante sarà l'organizzare i dati risultanti dalle effettive ripetizioni di tale esperimento.

Ad esempio, se l'esperimento fosse il lancio della moneta realizzato 10 volte, converrà prendere i 10 risultati e rappresentarli sinteticamente (cioè senza tener conto dell'ordine con cui sono usciti) con una tabella

$$X = \begin{cases} \text{testa} & \text{croce} \\ N_1 & N_2 \end{cases}$$

che ci dice appunto che testa è uscita N_1 volte, croce N_2: tra l'altro dovrà essere necessariamente $N_1 + N_2 = 10$.

Prima di generalizzare questo semplice esempio, notiamo che, come si è già detto per le variabili casuali, nella tabellina possiamo sempre sostituire agli attributi qualitativi testa e croce due valori numerici (valori argomentali) fissati per convenzione; ad esempio

$$\begin{array}{ll} \text{testa} & \rightarrow \quad x_1 = 0 \\ \text{croce} & \rightarrow \quad x_2 = 1 \end{array}$$

Tenendo presente anche tale possibilità, arriviamo a questa definizione generale:

> definiamo variabile statistica (a una dimensione) una tabella a due righe di valori numerici; gli elementi della prima riga sono detti valori argomentali, gli elementi della seconda, che devono essere tutti numeri interi, positivi o nulli, sono chiamati frequenze assolute
>
> $$X = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ N_1 & N_2 & \dots & N_n \end{array} \right. \tag{4.3}$$
>
> l'intero $N = \sum_{i=1}^{n} N_i$ rappresenta la numerosità degli individui (ovvero della popolazione) su cui la variabile statistica è stata costruita.

Si definisce frequenza relativa, o più semplicemente frequenza, di un valore argomentale x_i, il numero

$$f_i = \frac{N_i}{N} \ : \tag{4.4}$$

questo rappresenta la percentuale di individui, sulla popolazione totale, caratterizzati dal valore x_i. La variabile statistica può allora essere descritta con la forma

$$X = \left\{ \begin{array}{cc} x_1 & x_2 \dots x_n \\ f_1 & f_2 \dots f_n \ ; \end{array} \right. \tag{4.5}$$

alla (4.5) va poi aggiunto il valore N (numerosità della popolazione) così che dalla (4.5) si può risalire alla (4.3) e viceversa.

Notiamo che per la definizione (4.4) si ha ovviamente

$$\sum_{i=1}^{n} f_i = \frac{\sum_i N_i}{\sum_i N_i} = 1 \ ; \tag{4.6}$$

poiché inoltre è anche

$$f_i \geq 0$$

se ne deduce che la (4.5) definisce formalmente una distribuzione di probabilità concentrata sui valori $\{x_1, x_2 \ldots x_n\}$; basterà infatti porre

$$P(X = x_i) = f_i \quad . \tag{4.7}$$

Possiamo perciò dire che ogni definizione data, ogni proprietà dimostrata per le variabili casuali (in particolare per le variabili casuali discrete) deve necessariamente valere anche per le variabili statistiche, poiché formalmente queste sono identificabili con quelle per mezzo della (4.7).

Naturalmente tra i due tipi di variabili esiste una differenza sostanziale, di contenuto: mentre infatti nella variabile casuale i numeri p_i associati ai valori x_i misurano un grado di "possibilità" che il risultato dell'esperimento che stiamo descrivendo assuma il valore x_i, nella variabile statistica il numero f_i non fa che registrare il fatto che su N ripetizioni dell'esperimento si sono ottenuti N_i risultati con il valore x_i. La probabilità è un ente definito assiomaticamente, a priori, la frequenza è un indice che misura risultati empirici, definito a posteriori in base ad esperimenti già effettuati.

In base a questa identità formale possiamo ad esempio estendere alle variabili statistiche il concetto di funzione di distribuzione $F(x)$. Nel caso delle variabili statistiche $F(x)$ prende il nome di *funzione cumulativa di frequenza* e rappresenta la percentuale di elementi della popolazione, il cui valore argomentale x_i risulta minore o uguale ad x:

$$F(x) = \sum_{(i)} \ f_i = \frac{\sum_{(i)} N_i}{N} \ \text{ (per tutti gli } (i) \ \text{ per cui } x_i \leq x) \quad . \tag{4.8}$$

Per quanto già visto $F(x)$ risulta una funzione a gradini, crescente, compresa tra 0 e 1, con salti di altezza f_i in corrispondenza dei valori x_i.

5 La funzione densità di probabilità - Gli istogrammi delle variabili statistiche

a) Come si è già visto nel paragrafo precedente, una v.c. può essere caratterizzata per mezzo della sua funzione di distribuzione $F(x)$.

Se la v.c. X è continua, sarà particolarmente interessante, per una sua dettagliata descrizione, conoscere la probabilità che X stia in un intervallino del tipo $[x_0, x_0 + \Delta x]$; in base alla definizione di $F(x)$ [4] sarà

$$P(x_0 \leq X \leq x_0 + \Delta x) = F(x_0 + \Delta x) - F(x_0) \quad . \tag{5.1}$$

Per Δx abbastanza piccolo, si potrà allora porre, a meno di infinitesimi di ordine superiore e ammesso che $F(x)$ sia differenziabile,

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

$$f(x_0) = F'(x_0) = \lim_{\Delta x \to 0} \frac{P(x_0 \leq X \leq x_0 + \Delta x)}{\Delta x}.$$

La funzione $f(x)$ prende il nome di *densità di probabilità* ed essendo $F(x)$ una funzione monotona crescente, si avrà

$$f(x) \geq 0 \ , \quad \forall x \ . \tag{5.2}$$

Per comprendere meglio il significato di $f(x)$ si può ricorrere ad una analogia. Una distribuzione di probabilità infatti può essere assimilata

[4] Si noti che più esattamente sarà

$$F(x_0 + \Delta x) - F(x_0) = P(x_0 < X \leq x_0 + \Delta x) \ ;$$

poiché però X è per ipotesi continua, si ha anche $P(X = x_0) = 0$, e quindi

$$P(x_0 < X \leq x_0 + \Delta x) = P(x_0 \leq X \leq x_0 + \Delta x) \ .$$

ad una distribuzione di una massa (o di una carica) unitaria sull'asse reale; in questo caso $f(x)$ coinciderebbe col concetto di densità di massa (o di carica) sulla retta.

La funzione densità di probabilità è caratterizzante di una certa distribuzione di probabilità su R: da $f(x)$ infatti si può ricostruire $F(x)$ che, tenuto conto che $F(-\infty) = 0$, è data da

$$F(x) = \int_{-\infty}^{x} f(t)\ dt \quad . \tag{5.3}$$

Siccome poi si deve avere $F(+\infty) = 1$, si deduce che la $f(x)$ deve sempre soddisfare la condizione, detta di normalizzazione

$$1 = \int_{-\infty}^{+\infty} f(x)dx \quad . \tag{5.4}$$

Esempio 5.1: la densità di probabilità corrispondente all'Esempio 4.2 (distribuzione uniforme) è data da

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

$f(x)$ non è definito per $x = 0, 1$ dove $F(x)$ non è derivabile (punti angolosi).

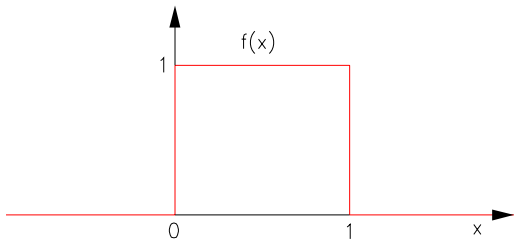


Figura 5.1

Esempio 5.2: la densità di probabilità corrispondente all'Esempio 4.3 è data da

$$f(x) = \frac{d}{dx}\left\{\frac{1}{\pi}\arctan x + \frac{1}{2}\right\} = \frac{1}{\pi}\frac{1}{1+x^2}$$

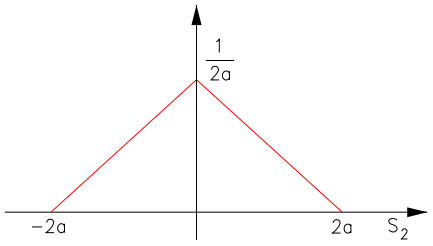


Figura 5.2

Si può notare che in base alla (5.3) si ha

$$\int_a^b f(x)dx = F(b) - F(a) = P(a \leq X \leq b) \tag{5.5}$$

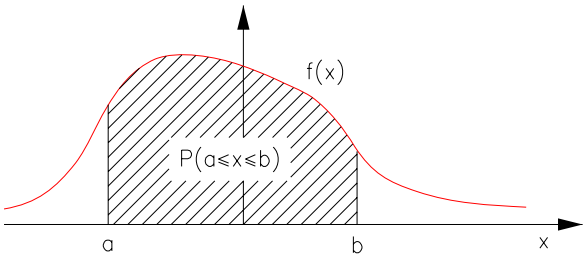


Figura 5.3

la (5.5) fornisce un legame diretto tra $f(x)$ e la distribuzione di probabilità della v.c da questa descritta. Più in generale, se A è un qualunque insieme misurabile sulla retta, si ha

$$P(X \in A) = \int_A f(x)dx \tag{5.6}$$

b) Il concetto di funzione densità di probabilità non è direttamente applicabile ad una variabile discreta poiché la sua funzione di distribuzione è in ogni punto o costante o discontinua: pertanto non si può formalmente definire un analogo della densità di probabilità per una v.s.[5]

E' tuttavia spesso importante fare un confronto tra una v.s. ed una v.c. descritta mediante la densità di probabilità. A questo scopo si può osservare in base all'esperienza che ripetendo N volte un esperimento il cui risultato sia descritto da una v.c continua, si ha una tendenza dei risultati a distribuirsi su tutti i possibili valori argomentali con un addensamento là dove $f(x)$ è maggiore .

Poiché un confronto (verifica) può essere fatto, in base a tutta la nostra impostazione, tra probabilità e frequenza, si potrà pensare ad esempio di fissare un intervallo $[x_0, x_0 + \Delta x]$ e di confrontare tra loro la $P(x_0 \leq X \leq x_0 + \Delta x)$ con la percentuale dei risultati che cadono nello stesso intervallo, ovvero

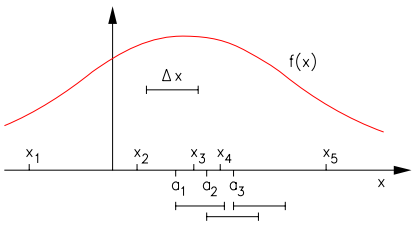
$$\begin{aligned} \Delta F(x_0) = & \quad \frac{N(x_0, \Delta x)}{N} \\ N(x_0, \Delta x) = & \quad \text{numero di elementi che hanno valore tra } x_0 \text{ e } x_0 + \Delta x. \end{aligned} \tag{5.7}$$

Naturalmente tale confronto è valido se N è abbastanza grande in modo che per tutti gli intervalli Δx di maggior interesse si abbia una numerosità sufficiente di elementi. Ad esempio è chiaro che con 5 estrazioni [6] da una variabile continua non si può dire nulla sulla frequenza in un intervallo del tipo Δx in figura; infatti basterebbero piccoli spostamenti di Δx per ottenere frequenze completamente diverse (fig. 5.4).

Più in generale si potranno esaminare simultaneamente sull'asse reale un certo numero di intervalli Δx_i: di solito questi sono scelti consecutivi uno all'altro e tali da ricoprire tutti i valori argomentali della v.s. Inoltre i Δx_i, che non sono necessariamente uguali fra loro, vanno scelti in modo che in ognuno di essi non cada un numero troppo piccolo di individui della v.s.

[5] Indichiamo con l'abbreviazione v.s. la variabile statistica.

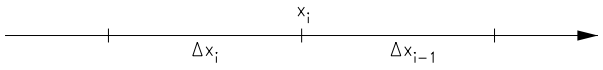
[6] Con "estrazione da una variabile casuale" indichiamo una ripetizione dell'esperimento i cui risultati sono descritti dalla variabile casuale stessa.



$$\Delta F(a_1) = \frac{N(a_1, \Delta x)}{5} = \frac{2}{5}$$
$$\Delta F(a_2) = \frac{N(a_2, \Delta x)}{5} = \frac{1}{5}$$
$$\Delta F(a_3) = \frac{N(a_3, \Delta x)}{5} = \frac{0}{5}$$

Figura 5.4

In genere, per evitare oscillazioni spurie delle frequenze, se un elemento ha come valore argomentale proprio un estremo x_i, tra due intervalli



l'elemento stesso viene contato per una metà in Δx_{i-1} e per l'altra metà in Δx_i

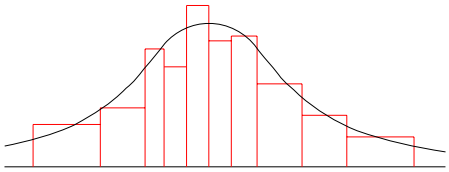
Per ottenere un più facile confronto grafico, notiamo che per la definizione di densità di probabilità, se gli intervalli Δx_i non sono troppo grandi si ha

$$(x \in \Delta x_i) \ , \ f(x) \cong \frac{P(X \in \Delta x_i)}{\Delta x_i} \tag{5.8}$$

risulta perciò spontaneo confrontare il valore di $f(x)$ in Δx_i con il valore

$$h_i = \frac{\Delta F(x_i)}{\Delta x_i} = \frac{N(x_i, \Delta x_i)}{N \cdot \Delta x_i}. \tag{5.9}$$

La funzione a gradini che ha in ogni intervallino Δx_i il valore h_i prende il nome di istogramma delle v.s.



Tra l'istogramma di una v.s., generata ripetendo N volte un esperimento, e la densità di probabilità relativa ad esso, ci si aspetta un rapporto del tipo in figura

Figura 5.5

Esempio 5.3: si consideri una v.c. Z distribuita secondo la densità di probabilità

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \quad ; \tag{5.10}$$

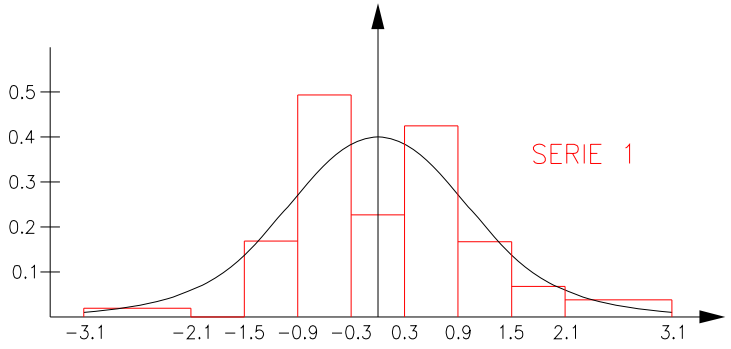
una tale variabile prende il nome di *variabile normale standardizzata*.

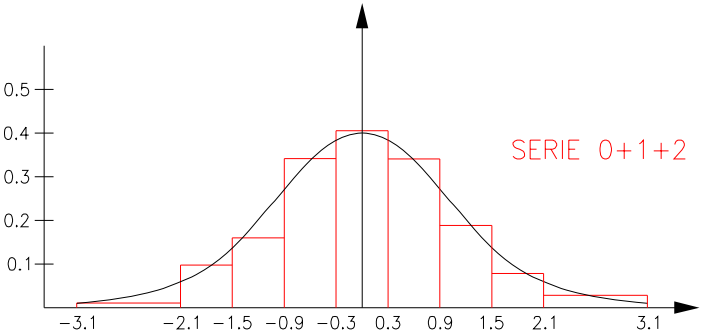
	0	1	2		0	1	2
00	-0.179	-0.399	-0.235	25	-0.927	0.838	-1.546
01	0.421	1.454	0.904	26	1.232	2.170	0.088
02	0.210	-0.556	0.465	27	0.935	0.665	2.034
03	-1.598	0.019	-0.266	28	-1.739	-0.622	-1.563
04	1.717	1.514	-0.012	29	0.990	-1.483	0.154
05	-0.308	0.867	-0.372	30	-0.189	-0.240	0.133
06	-0.421	0.516	-0.038	31	1.866	-1.398	0.068
07	-0.776	0.874	-1.265	32	-0.018	0.628	0.230
08	0.640	-0.522	0.023	33	-0.646	-0.350	0.324
09	-0.319	0.889	1.180	34	-1.150	-0.220	-0.533
10	0.610	-0.383	1.812	35	0.103	-0.361	1.024
11	-0.174	-0.154	0.098	36	1.243	0.539	0.684
12	2.576	-0.684	-1.200	37	-0.093	-1.190	0.580
13	-1.103	1.398	-0.653	38	-0.261	-0.194	0.303
14	1.635	0.448	-1.530	39	-0.230	-0.550	0.266
15	-0.068	-0.860	-0.194	40	-0.148	0.504	-0.028
16	-1.960	1.076	-0.671	41	1.122	0.896	-0.789
17	0.443	-0.912	0.251	42	0.499	-1.032	0.159
18	1.360	0.533	1.094	43	0.678	-0.782	0.470
19	0.810	0.319	-1.514	44	-1.347	3.090	-0.896
20	0.616	1.347	-1.866	45	-1.094	-0.610	-0.287

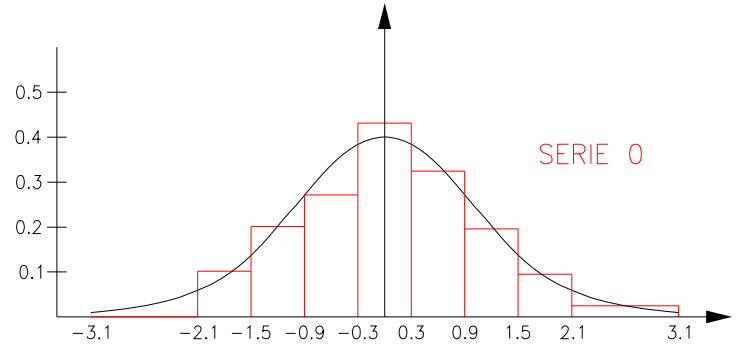
21	-0.598	-2.366	-0.831	46	-0.088	-0.889	0.803
22	0.426	1.580	-1.112	47	0.093	-0.476	1.265
23	0.831	-0.516	-1.717	48	-0.950	-0.008	0.012
24	-0.640	-0.128	1.276	49	-0.329	-0.138	-0.504

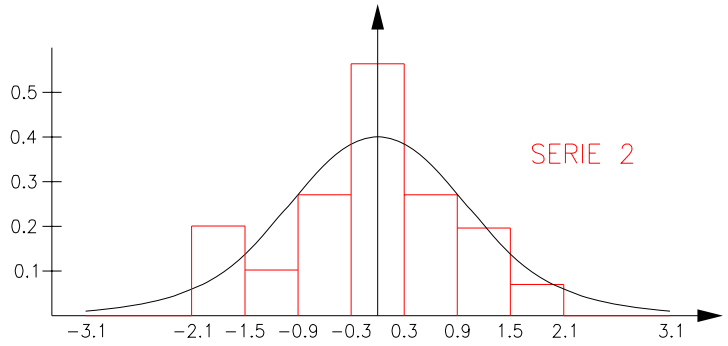
Tabella 1: Tavola di 150 valori estratti dalla normale standardizzata.

I valori della $f(x)$ per diversi valori di x si trovano nella Tabella 5.1. Compiendo 150 estrazioni da questa variabile otteniamo i valori, arrotondati alla terza cifra, riportati nella tabella. Dividendo i 150 risultati


0.5
0.4
0.3
0.2
0.1
SERIE 1
−3.1
−2.1
−1.5
−0.9
−0.3
0.3
0.9
1.5
2.1
3.1


0.5
0.4
0.3
0.2
0.1
SERIE 0+1+2
−3.1
−2.1
−1.5
−0.9
−0.3
0.3
0.9
1.5
2.1
3.1


0.5
0.4
0.3
0.2
0.1
SERIE 0
−3.1
−2.1
−1.5
−0.9
−0.3
0.3
0.9
1.5
2.1
3.1


0.5
0.4
0.3
0.2
0.1
SERIE 2
−3.1
−2.1
−1.5
−0.9
−0.3
0.3
0.9
1.5
2.1
3.1

ottenuti in tre gruppi da 50, vogliamo tracciare l'istogramma delle tre v.s. confrontandolo con la curva teorica: ripeteremo poi la stessa cosa riunendo tutti i 150 risultati in una sola variabile statistica.

Intervalli			Frequenze relative				p
			serie 0	serie 1	serie 2	serie 0+1+2	
-3.1	:	-2.1	0.00	0.02	0.00	0.006	0.017
-2.1	:	-1.5	0.06	0.00	0.12	0.060	0.049
-1.5	:	-0.9	0.12	0.10	0.06	0.093	0.117
-0.9	:	-0.3	0.16	0.30	0.16	0.207	0.198
-0.3	:	0.3	0.26	0.16	0.36	0.260	0.236
0.3	:	0.9	0.20	0.26	0.14	0.200	0.198
0.9	:	1.5	0.12	0.18	0.12	0.107	0.117
1.5	:	2.1	0.06	0.04	0.04	0.047	0.049
2.1	:	3.1	0.02	0.04	0.00	0.020	0.017

Tab. 5.2

In base ai valori estratti si è ravvisata l'opportunità di costruire l'istogramma sui 9 intervalli della Tabella 5.2, nella quale, oltre alle frequenza relative agli intervalli Δx_i per ciascuna serie, sono anche riportati i valori teorici $P(x \in \Delta x_i)$ per la variabile normale standardizzata, ricavati dalle tabelle allegate alla Parte I. Dalla Tabella 5.2 è facile calcolare le ordinate degli istogrammi riportati in figura 5.6. Si noti che tanto nel confronto tra gli istogrammi, quanto dal confronto della 5^a e 6^a colonna di Tabella 5.2 risulta evidente come il campione complessivo, di 150 elementi, è distribuito in modo assai più vicino alla curva teorica di quanto non lo siano i campioni di 50 elementi.

6 Variabile casuale funzione di un'altra

Indichiamo con X la v.c. che rappresenta il lancio di un dado non truccato: poiché chiaramente

$$P\{X \in \text{ (pari)}\} = 1/2$$
$$P\{X \in \text{ (dispari)}\} = 1/2$$

$$\{x \text{ (pari)}\} \cup \{x \text{ (dispari)}\} = S \equiv \{1,2,3,4,5,6\}$$
$$\{x \text{ (pari)}\} \cap \{x \text{ (dispari)}\} = \emptyset$$

facendo uso della seguente corrispondenza

$$y = g(x)\ ; \quad \begin{array}{ll} x \text{ pari} & \to y \text{ testa} \\ x \text{ dispari} & \to y \text{ croce} \end{array}$$

si arriva a costruire una nuova v.c. Y

$$Y = \left\{ \begin{array}{cc} \text{testa} & \text{croce} \\ 1/2 & 1/2 \end{array} \right.$$

formalmente identica a quella che descrive il lancio di una moneta non truccata. Il punto essenziale di questo esempio sta nel fatto che si è usata la corrispondenza $y = g(x)$ per definire una nuova distribuzione di probabilità sui valori di Y: la corrispondenza è illustrata nella figura 6.1.

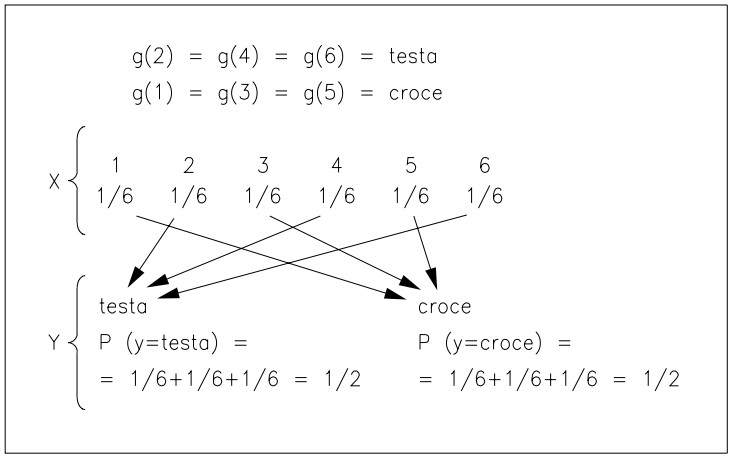


Figura 6.1

Questo procedimento può essere generalizzato anche al caso continuo, in cui una funzione numerica

$$y = g(x)$$

sia definita sull'insieme S_X (insieme dei valori argomentali della X).

La funzione g trasformerà S_X nella corrispondente immagine S_Y. Sia ora A_Y un sottoinsieme di S_Y, vi sarà in corrispondenza un $A_X (\subseteq S_X)$ tale che

$$g(A_X) = A_Y \quad ;$$

l'insieme A_X si chiama immagine inversa di A_Y e se, tutte le volte che A_Y è misurabile in senso ordinario (misura di Lebesgue sulla retta) lo é anche A_X, diciamo che la funzione g stessa è misurabile.

Supponendo g misurabile, porremo per definizione

$$P(Y \in A_Y) = P(X \in A_X) \ . \tag{6.1}$$

In figura 6.2 ad esempio si ha $A_Y = \{c \leq y \leq d\}$, così che risulta

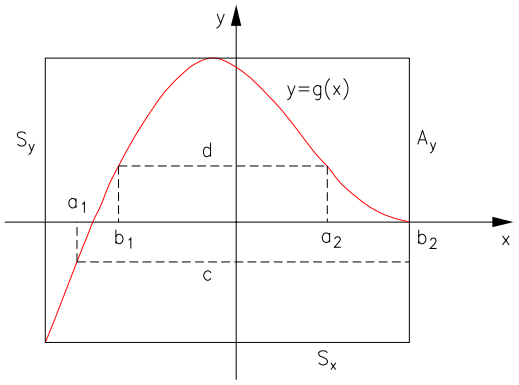


Figura 6.2

$$A_X = \{a_1 \leq x \leq b_1\} \cup \{a_2 \leq x \leq b_2\} \quad .$$

Dalla (6.1) è poi semplice derivare la densità di probabilità della Y una volta nota quella della X ed assumendo una qualche maggior regolarità

per $g(x)$, ad esempio che tale funzione sia differenziabile. Indicheremo con $f_Y(y)$ la densità di probabilità di Y e con $f_X(x)$ quella di X.

Sia allora $A_Y = dy(y_0)$ un intervallino attorno ad y_0 e di ampiezza dy. Se, come supporremo, $g(x)$ è una funzione continua e differenziabile, A_X sarà un insieme formato da uno o più intervallini

$$A_X = \bigcup_i dx_i(x_i)$$

attorno ai punti x_i tali che

$$g(x_i) \cong y_0$$

(fig. 6.3).

In base alla (6.1) si avrà

$$P(Y \in dy(y_0)) = \sum_i P\left(X \in dx_i(x_i)\right) \tag{6.2}$$

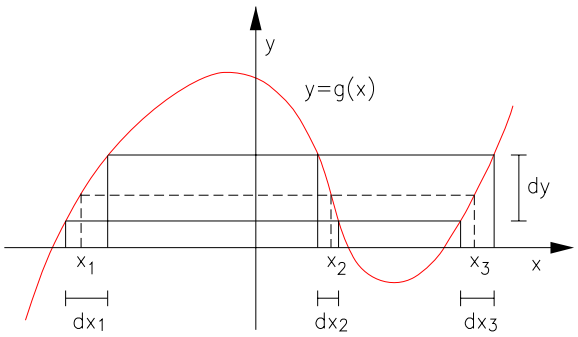


Figura 6.3

Ora notiamo che, per definizione, per una variabile X qualunque, tanto $P(X \in A)$ quanto $f_X(x)$ devono essere numeri positivi o al più nulli, così che volendo riscrivere la (5.6) per un intervallo infinitesimo $dx(x)$ (attorno al punto x) si dovrà porre

$$P(X \in dx(x)) = f_X(x)|dx| \quad [7] \tag{6.3}$$

dove con $|dx|$ intendiamo la misura dell'intervallino presa, naturalmente, con segno positivo.

Fatta questa precisazione, dividiamo la (6.2) per $|dy|$ ottenendo, con ovvio passaggio, la relazione

$$\frac{P(Y \in dy(y_0))}{|dy|} = \sum_i \frac{P(X \in dx_i(x_i))}{|dx_i|} \frac{1}{\left|\frac{dy}{dx_i}\right|}$$

così che ricordando la (6.3) si trova infine

$$f_Y(y_0) = \sum_i \frac{f_X(x_i)}{\left|\frac{dy}{dx_i}\right|} = \sum_i \frac{f_X(x_i)}{|g'(x_i)|} \quad . \tag{6.4}$$

Esempio 6.1: sia

$$y = ax + b \ ; \tag{6.5}$$

qualunque sia $f_X(x)$ si trova per $f_Y(y)$

$$f_Y(y) = \frac{f_X(x)}{|a|}$$

dove x e y sono legati dalla (6.5); ovvero

$$f_Y(y) = \frac{f_X(\frac{y-b}{a})}{|a|} \quad . \tag{6.6}$$

Così, se prendiamo $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, si trova

[7] L'uguaglianza va intesa, nel senso degli infinitesimi, al primo ordine in dx.

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|a|} e^{\frac{-(y-b)^2}{2a^2}} \quad . \tag{6.7}$$

Osservazione 6.1: si può osservare che nel caso la $g(x)$ sia una funzione monotona in senso stretto (crescente o decrescente), la funzione inversa $x = g^{(-1)}(y)$ avrà una sola determinazione, così che la somma in (6.4) si riduce ad un solo elemento, inoltre in questo caso $|\frac{dy}{dx}| = \pm g'(x)$ a seconda che g sia crescente o decrescente.

Esempio 6.2: sia $y = x^2$ ed $f_X(x)$ qualunque. Dato y, l'equazione $x^2 = y$ ha le due radici (fig. 6.4): $x_1 = -\sqrt{y}$, $x_2 = \sqrt{y}$

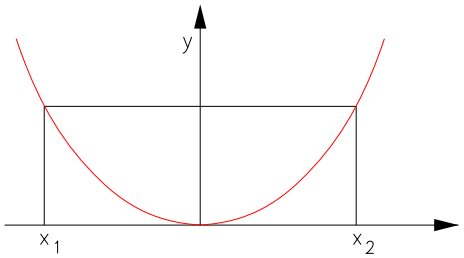


Figura 6.4

In corrispondenza si ha

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

Dalla (6.4) allora si trova

$$f_Y(y) = \frac{f_X(-\sqrt{y})}{|-2\sqrt{y}|} + \frac{f_X(\sqrt{y})}{|2\sqrt{y}|} = \frac{f_X(-\sqrt{y}) + f_X(\sqrt{y})}{2\sqrt{y}} \quad . \tag{6.8}$$

Se poi in particolare si pone $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, si ottiene

$$f_Y(y) = \frac{e^{-y/2}}{\sqrt{2\pi}\sqrt{y}} \quad (y \geq 0) \ . \tag{6.9}$$

Osservazione 6.2: si noti che se X è una qualsiasi v.c. con densità $f_X(x)$ e una funzione di distribuzione $F_X(x)$, è possibile costruire una nuova v.c. Y definita da

$$Y = F_X(X) \ . \tag{6.10}$$

Quando x percorre l'insieme di definizione di $X, y = F_X(x)$ varia sull'intervallo [0,1].

Al solito si potrà poi estendere la Y a tutto l'asse reale, assegnando probabilità zero al complemento di [0,1].

Calcoliamo ora la densità di Y, tenendo conto che $F_X(x)$ è monotona crescente; si ha

$$f_Y(y) = \frac{f_X(x)}{\frac{d}{dx}F_X(x)} = 1 \ , \quad y \in [0,1] \tag{6.11}$$

cioè Y è uniformemente distribuita su [0,1].

Reciprocamente si prova che se U è una v.c. uniformemente distribuita su [0,1] e $g(u)$ una funzione monotona crescente

$$X = g(U) \tag{6.12}$$

é una v.c con funzione di distribuzione

$$F_X(x) = g^{(-1)}(x). \tag{6.13}$$

Questa osservazione è di grande utilità per la generazione di variabili statistiche estratte da v.c. con distribuzione assegnata qualunque. In effetti é facile ottenere, tramite un computer, estrazioni dalla distribuzione uniforme: usando le (6.12), (6.13) queste possono essere trasformate in estrazioni da una variabile qualsivoglia. Questa possibilità é alla base della costruzione di simulazioni di esperimenti stocastici.

7 La media

Nei paragrafi precedenti abbiamo visto come sia possibile caratterizzare una v.c. per mezzo di funzioni che permettono di costruire la distribuzione di probabilità della variabile stessa.

E' chiaro che una completa conoscenza di una v.c. monodimensionale, considerata isolatamente, può derivare solo dalla sua funzione di distribuzione o da ogni altra informazione equivalente a questa. Tuttavia, per molte distribuzioni di interesse pratico, la variabile è piuttosto ben localizzata, nel senso che esiste un intervallo di piccole proporzioni su cui si ha un'alta probabilità: ad esempio è oggi semplice, con opportuni strumenti, misurare distanze di 1 km, ottenendo misure che differiscono tra loro per 2-3 mm al più.

Per variabili di questo tipo le informazioni di gran lunga più importanti possono essere riassunte rispondendo alle due domande: dove è localizzata la distribuzione? (nell'esempio della misura della distanza D avremmo potuto rispondere che D è localizzata attorno a 1 km), quanto è dispersa la distribuzione? (nell'esempio si poteva dire che pressoché il 100% di probabilità di D era concentrato in un intervallo di 3 mm di lunghezza). Per dare una risposta più precisa a tali domande introduciamo i concetti di media e varianza che sono i due indici, caratteristici di una distribuzione, più importanti.

Per definizione si chiama media della variabile X continua, con densità $f(x)$, il numero (quando esiste)

$$E\{X\} = \int_{-\infty}^{+\infty} x\ f(x)\ dx \quad ; \tag{7.1}$$

nel caso di una v.c. X discreta si ha per definizione

$$E\{X\} = \sum_i x_i p_i \tag{7.2}$$

dove la sommatoria corre su tutti i valori argomentali; per una v.s. si pone, in analogia alla (7.2),

$$E\{X\} = \sum_i x_i f_i = \frac{\sum_i x_i N_i}{N} \tag{7.3}$$

La media di una v.c. X verrà indicata come $\mu(X), \mu_X$ o semplicemente μ se non vi sia ambiguità; per una v.s. X invece si porrà $m(X), m_X$ o semplicemente m.

Con $E\{\bullet\}$ invece intendiamo l'operazione matematica che, data una distribuzione, sia essa a priori (v.c.) o a posteriori (v.s.), permette di calcolare un numero, appunto la media della distribuzione stessa.

Esempio 7.1: come accennato si può parlare di media di una v.c. quando i numeri dati da (7.1) o da (7.2) sono ben definiti.

In effetti vi sono variabili casuali sia continue che discrete che non ammettono media. Ad esempio la v.c. di Cauchy, già presentata nell'Esempio 5.2, con densità di probabilità

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

non ammette media in quanto la funzione $\frac{1}{\pi}\frac{1}{1+x^2}$ non é assolutamente integrabile su tutto l'asse.

Analogamente una v.c. discreta con valori argomentali $X_k = k, (k = 1, 2 \ldots)$, e con distribuzione

$$P\{X = k\} = p_k = \frac{6}{\pi^2}\frac{1}{k^2} \ , \ (k = 1, 2 \ldots)$$

non ammette media in quanto risulta

$$\sum_{k=1}^{+\infty} x_k p_k = \frac{6}{\pi^2}\sum_{k=1}^{+\infty}\frac{1}{k} = +\infty \ .$$

Osservazione 7.1: se interpretiamo la distribuzione di probabilità come una distribuzione di una massa unitaria sull'asse x, allora il concetto di media coincide con quello di baricentro.

Osservazione 7.2: se una distribuzione di probabilità è simmetrica rispetto ad un valore c, nel senso che, ad esempio,

$$f(c+h) = f(c-h) \qquad \forall h \ ,$$

allora c coincide con la media

$$E\{X\} = c \ ;$$

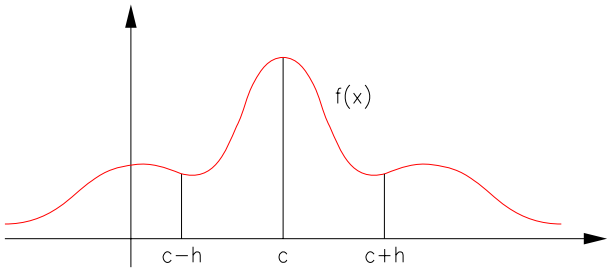


Figura 7.1

infatti

$$\begin{aligned} E\{X\} &= \int_{-\infty}^{+\infty} xg(x)dx = \int_{-\infty}^{+\infty} (c+h)f(c+h)dh = \\ &= c\int_{-\infty}^{+\infty} f(c+h)dh + \int_{-\infty}^{+\infty} hf(c+h)dh = c \quad . \end{aligned}$$

Osservazione 7.3: poiché per una distribuzione continua si ha

$$dP(x) = dF(x) = f(x)dx$$

la (7.1) può anche essere scritta come

$$E\{X\} = \int_{-\infty}^{+\infty} xdP(x) = \int_{-\infty}^{+\infty} xdF(x) \quad : \tag{7.4}$$

la (7.4) è generalizzabile in modo tale da comprendere anche il caso discreto: perciò la (7.4) è un modo di rappresentare la media che comprende sia la (7.1) che la (7.2).

Osservazione 7.4: nel caso di una v.s., dalla (7.3) si può scrivere

$$m_X = \frac{1}{N}\sum_i x_i N_i \tag{7.5}$$

notando che N_i è il numero di volte che il valore argomentale x_i è risultato estratto, la (7.5) dice che la media della v.s. è la somma di tutti i valori

della X sulla popolazione che la compone, divisa per la numerosità della popolazione stessa.

Se con x_i indichiamo in sostanza ogni singolo valore estratto anziché il valore argomentale, si potrà supporre

$$m_X = \frac{1}{N} \sum_i x_i \quad [8] \tag{7.6}$$

La (7.6) è di particolare utilità pratica quando la v.s. non sia ordinata per valori argomentali e sia di numerosità elevata.

Consideriamo ora una variabile Y funzione di una X nota, secondo la legge

$$Y = g(X) \quad .$$

Sappiamo che per definizione di media si può supporre

$$E\{Y\} = \int_{-\infty}^{+\infty} y f_Y(y) dy$$

e sappiamo anche che tramite la (6.4) è possibile ricavare la $f_Y(y)$ a partire dalla $f_X(x)$ nota: tuttavia l'applicazione della (6.4) spesso non è semplice, mentre sarebbe desiderabile poter calcolare μ_Y direttamente dalla $f_X(x)$. Questo problema è risolto dall'importante *teorema della media.*

Indichiamo con $E_Y\{\bullet\}$ ed $E_X\{\bullet\}$ le operazioni

$$E_Y\{\bullet\} = \int_{-\infty}^{+\infty} \bullet f_Y(y) dy \ , \ E_X\{\bullet\} = \int_{-\infty}^{+\infty} \bullet f_X(x) dx \ ;$$

[8] Data ad esempio la variabile (2,3,2,5,3) risulta perfettamente equivalente porre

$$m_X = \frac{1}{5}(2+3+2+5+3) = 3$$

oppure

$$m_X = 2 \cdot \frac{2}{5} + 3 \cdot \frac{2}{5} + 5 \cdot \frac{1}{5} = 3 \ .$$

in coerenza con questa notazione definiamo

$$E_X\{g(X)\} = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

notando che questa definizione è puramente analitica e per ora non ha alcuna relazione con le v.c. X e Y.

Premesso ciò, si ha il seguente teorema.

> **Teorema della media.** Se le due variabili casuali X e Y sono legate dalla relazione $Y = g(X)$ allora la media Y, se esiste, è data da
>
> $$\mu_Y = E_Y\{Y\} = E_X\{g(X)\} \ . \qquad (7.7)$$

In pratica la (7.7) ci dice che è possibile fare il cambiamento di variabili nell'operazione di media, così che d'ora in poi sarà possibile tralasciare (tranne che in casi particolari) l'indicazione della variabile cui l'operatore $E\{\bullet\}$ si riferisce.

Dimostriamo il teorema nel caso semplice in cui X sia una v.c. continua e $y = g(x)$ sia una funzione monotona, strettamente crescente: in tal caso infatti ($g'(x) > 0$)

$$f_Y(y) = \frac{f_X(x)}{g'(x)} \quad \text{ove } x = g^{(-1)}(y)$$

così che

$$E_Y\{Y\} = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_{-\infty}^{+\infty} y \frac{f_X(x)}{g'(x)} \, dy$$

Se ora nel fare l'integrazione operiamo la sostituzione di variabile $y = g(x)$ (x = nuova variabile di integrazione), allora si trova

$$\begin{aligned} E_Y\{y\} &= \int_{-\infty}^{+\infty} g(x) \frac{f_X(x)}{g'(x)} \, g'(x) dx = \int_{-\infty}^{+\infty} g(x) f_X(x) dx = \\ &= E_X\{g(X)\} \end{aligned}$$

come volevasi dimostrare.

Esempio 7.2: sia

$$f_X(x) = \begin{cases} \frac{2}{\pi} & 0 \le x \le \frac{\pi}{2} \\ 0 & x < 0,\ x > \frac{\pi}{2} \end{cases}$$

e sia $y = \sin x$.

In questo caso $0 \le x \le \frac{\pi}{2}$ viene trasformato in $0 \le y \le 1$, così che si ha

$$f_Y(y) = \frac{\frac{2}{\pi}}{\cos x} = \frac{2}{\pi\sqrt{1-y^2}} \quad (0 \le y \le 1) \quad .$$

Pertanto

$$E_Y\{Y\} = \frac{2}{\pi}\int_0^1 \frac{y\,dy}{\sqrt{1-y^2}} = \frac{2}{\pi}[-\sqrt{1-y^2}]_0^1 = \frac{2}{\pi} \quad .$$

Verifichiamo che, in base al teorema della media,

$$E_Y\{Y\} = E_X\{g(X)\}\ ;$$

infatti

$$E_X\{g(X)\} = \frac{2}{\pi}\int_0^{\pi/2} \sin\ x\ dx = \frac{2}{\pi}[-\cos x]_0^{\pi/2} = \frac{2}{\pi}$$

c.v.d.

Per il caso più generale in cui $g(x)$ non è monotona, anziché fare una dimostrazione generale ci avvaliamo del seguente esempio semplice.

Esempio 7.3: sia

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

e sia $Y = X^2$: poiché questa non è una funzione monotona sull'asse x, occorre vedere con attenzione come verificare il teorema della media.

Ricordiamo in primo luogo (Esempio 6.2) che la Y ha una densità di probabilità data da

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \frac{e^{-y/2}}{\sqrt{y}} \quad (y \geq 0) \ ,$$

così che

$$E\{Y\} = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} y \frac{e^{-y/2}}{\sqrt{y}} \, dy = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \sqrt{y} e^{-y/2} \, dy \ .$$

Integrando una volta per parti ed usando il fatto che

$$\frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \frac{e^{-y/2}}{\sqrt{y}} \, dy = 1 \ ,$$

si trova facilmente

$$\mu_Y = 1 \ .$$

Ora ricordiamo che si può porre

$$E\{Y\} = \int_0^{+\infty} y \, dP(y) \quad ,$$

ed osserviamo che ogni elemento $dP(y)$ deriva da due componenti $dP(x_1)$, $dP(x_2)$ come in fig. 7.2

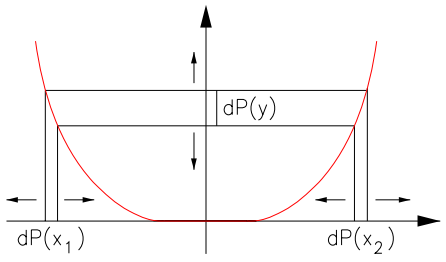


Figura 7.2

$$dP(y) = dP(x_1) + dP(x_2) \quad .$$

Inoltre quando y si muove da 0 a $+\infty$, corrispondentemente x_1 si muove da 0 a $-\infty$, ed x_2 da 0 a $+\infty$; così che si ha

$$\begin{aligned} E_Y\{Y\} &= \int_0^{+\infty} y\ dP(y) = \int_{-\infty}^{0} y\ dP(x_1) + \int_0^{+\infty} y\ dP(x_2) = \\ &= \int_{-\infty}^{0} x_1^2 dP(x_1) + \int_0^{+\infty} x_2^2 dP(x_2) \\ &= \int_{-\infty}^{+\infty} x^2 dP(x) = E_X\{g(X)\} \quad . \end{aligned}$$

Per verificare ciò nel nostro esempio occorre calcolare

$$\begin{aligned} E_X\{g(X)\} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} [-x e^{-x^2/2}]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = 1 \ , \end{aligned}$$

c.v.d.

Dal teorema della media discendono tre semplici ma importantissimi corollari.

Corollario 7.1: *L'operazione di media è un'operazione lineare.* Infatti se $Y = aX + b$, allora è anche

$$\begin{aligned} E_Y\{Y\} &= E_X\{aX + b\} = \int_{-\infty}^{+\infty} (ax + b) f_X(x) dx = \qquad (7.8) \\ &= a \int_{-\infty}^{+\infty} x f_X(x) dx + b \int_{-\infty}^{+\infty} f_X(x) dx = aE\{X\} + b \end{aligned}$$

In particolare questo comporta che, data una X qualunque, la trasformazione

$$Y = X - \mu_X$$

genera una nuova variabile, detta scarto della X dalla media, che ha media nulla:

$$E\{Y\} = E\{X\} - \mu_X = 0 \ .$$

Corollario 7.2: sia $Y = g(X)$. Sotto opportune ipotesi si dimostra che, con una certa approssimazione, vale

$$\mu_Y \cong g(\mu_X) \tag{7.9}$$

formula questa di evidente utilità pratica.

Sia X una variabile concentrata in una zona dell'asse x, cioè con una alta probabilità su un intervallo relativamente piccolo; supponiamo inoltre che in corrispondenza a tale zona la $g(x)$ abbia un andamento molto regolare, senza rapide variazioni (fig. 7.3).

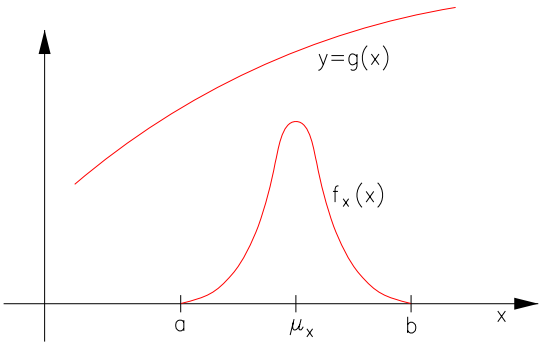


Figura 7.3

Indichiamo con $[a, b]$ l'intervallo attorno alla media μ_X, su cui entrambe le condizioni sono soddisfatte.

Su $[a, b]$ si può porre con buona approssimazione

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

così che si ha

$$\begin{aligned}
\mu_Y = E\{Y\} &= \int_{-\infty}^{+\infty} g(x) f_X(x) dx \cong \\
&= \int_a^b [g(\mu_X) + g'(\mu_X)(x - \mu_X)] f_X(x) dx = \\
&= g(\mu_X) \int_a^b f_X(x) dx + g'(\mu_X) \int_a^b (x - \mu_X) f_X(x) dx \cong \\
&\cong g(\mu_X) \int_{-\infty}^{+\infty} f_X(x) dx + g'(\mu_X) \int_{-\infty}^{+\infty} (x - \mu_X) f_X(x) dx = \\
&= g(\mu_X) \ ,
\end{aligned}$$

dove si è tenuto conto che

$$\begin{aligned}
&\int_{-\infty}^{+\infty} f_X(x) dx = 1 \\
&\int_{-\infty}^{+\infty} (x - \mu_X) f_X(x) dx = 0 \ .
\end{aligned}$$

Corollario 7.3: l'operazione di media è positiva nel senso che se $g(x) \geq 0$, allora $E\{g(X)\} \geq 0$.
In effetti se g è una funzione a valori non negativi, posto $Y = g(X)$ sarà $P\{Y < 0\} = 0$ e quindi anche

$$\mu_Y = E\{Y\} = E\{g(X)\} \geq 0 \ .$$

Una conseguenza di questo corollario è che se sono date due funzioni g, h tali che

$$g(x) \leq h(x)$$

allora è anche

$$E\{g(X)\} \leq E\{h(X)\} \ , \tag{7.10}$$

ammesso che tali quantità siano definite.
Infatti da

$$0 \leq E\{h(X) - g(X)\} = E\{h(X)\} - E\{g(X)\} \ ,$$

discende immediatamente la (7.10).

8 La varianza - Il teorema di Tchebycheff

Come già detto nel paragrafo precedente, la varianza è un indice che misura il grado di concentrazione di una v.c. X attorno alla media.

Si pone per definizione, ammesso che esista,

$$\sigma^2(X) = E\{(X - \mu_X)^2\} \ . \tag{8.1}$$

La precisazione non è banale; si consideri, ad esempio la distribuzione di Cauchy

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2} \ ,$$

che, come già detto, non ha media, anche a fortiori non può ammettere varianza; più in generale una $f(x)$ per cui $f(x) = 0(|x|^{-\alpha})$ con $2 < \alpha \leq 3$ è una densità di probabilità che non ammette varianza, benché esista la media.

Indicheremo la varianza di una v.c. con $\sigma^2(X), \sigma_X^2$ o semplicemente σ^2, mentre per una v.s. useremo i simboli $S^2(X)$, S_X^2 o semplicemente S^2.

La radice quadrata della varianza, σ, o rispettivamente S, viene chiamato scarto quadratico medio (abbreviato s.q.m) ed è pure un indice molto usato perché è dimensionalmente omogeneo alla variabile X.

In forma esplicita la (8.1) può essere riscritta come

$$\sigma^2(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx \tag{8.2}$$

$$\sigma^2(X) = \sum_i (x_i - \mu_X)^2 p_i \tag{8.3}$$

$$S^2(X) = \sum_i (x_i - m_X)^2 \frac{N_i}{N} \tag{8.4}$$

$$S^2(X) = \frac{1}{N} \sum_i (x_i - m_X)^2 \tag{8.5}$$

secondo che ci si riferisca ad una v.c. continua o discreta [(8.2) e (8.3)] o ad una v.s. ordinata per valori argomentali o non ordinata, ovvero con ripetizioni [(8.4) e (8.5)].

Dalla (8.1), sviluppando il quadrato ed usando la linearità della media, si trova

$$\begin{aligned} \sigma^2(X) &= E\{X^2 - 2\mu_X X + \mu_X^2\} = \\ &= E\{X^2\} - 2\mu_X E\{X\} + \mu_X^2 = \\ &= E\{X^2\} - \mu_X^2 \quad . \end{aligned} \tag{8.6}$$

La (8.6) che, nel caso di una v.s. non ordinata si scrive

$$S^2(X) = \frac{1}{N} \sum_i x_i^2 - m_X^2 \ ,$$

é particolarmente utile nei conti pratici perché evita di calcolare tutte le differenze (scarti) $x_i - m_X$.

La quantità $E\{X^2\}$ si chiama momento del secondo ordine della X; la (8.6) può anche essere letta dicendo che il momento del secondo ordine della X è dato dal quadrato della media più la varianza.

Notiamo che nell'analogia meccanica in cui la probabilità viene considerata come una distribuzione di una massa unitaria sull'asse delle x, la varianza assume il senso di momento di inerzia della distribuzione rispetto al baricentro.

Anche in rapporto a tale analogia risulta chiaro che più le masse sono dislocate lontano dal baricentro, più alto è il momento di inerzia; ovvero, meno la probabilità è concentrata attorno alla media, più alta é la varianza e viceversa.

Per rendere più precisa tale nozione dimostriamo il teorema di Tchebycheff, che ha il grande pregio di essere valido per una distribuzione qualunque.

Teorema di Tchebycheff. Preso un qualunque numero $\lambda > 1$, per una qualunque variabile casuale X, che ammette media e varianza finite, vale la diseguaglianza

$$P(|X - \mu_X| \leq \lambda\sigma_X) \geq 1 - \frac{1}{\lambda^2} \tag{8.7}$$

La (8.7) dice che più piccolo è lo s.q.m. σ_X, più piccolo é l'intervallo $[\mu_X - \lambda\sigma_X, \mu_X + \lambda\sigma_X]$ in cui si é sicuri (qualunque sia X!) di avere una probabilità che al minimo vale $1 - \frac{1}{\lambda^2}$.

Così ad esempio:

in un intervallo $\mu_X - 2\sigma_X, \mu_X + 2\sigma_X$ si ha almeno il 75% di probabilità;
in un intervallo $\mu_X - 3\sigma_X, \mu_X + 3\sigma_X$ si ha almeno l'89% di probabilità.

Per dimostrare la (8.7), consideriamo la funzione

$$h_\lambda(x) = \begin{cases} \lambda^2\sigma^2 & ; \quad |x - \mu| > \lambda\sigma \\ 0 & ; \quad |x - \mu| \leq \lambda\sigma \ . \end{cases} \tag{8.8}$$

Posto

$$P_0 = P\{|x - \mu| \leq \lambda\sigma\} \quad P_\lambda = 1 - P_0 = P\{|x - \mu| > \lambda\sigma\}$$

si vede che la v.c. $h_\lambda(x)$ è discreta e che

$$h_\lambda(X) = \begin{cases} 0 & \lambda^2\sigma^2 \\ P_0 & P_\lambda \ , \end{cases}$$

così che

$$E\{h_\lambda(X)\} = \lambda^2\sigma^2 P_\lambda \ .$$

Ora è chiaro dalla definizione (8.8) che

$$h_\lambda(x) \leq (x-\mu)^2 \ , \ \forall x$$

così che applicando la (7.10) si ha

$$\sigma^2 = E\{(X-\mu)^2\} \geq E\{h_\lambda(X)\} = \lambda^2\sigma^2 P_\lambda$$

e quindi $P_\lambda \leq 1/\lambda^2$, ovvero $P_0 \geq 1 - 1/\lambda^2$, che coincide con la (8.7).

Notiamo che il teorema di Tchebycheff giustifica pienamente l'affermazione che media e varianza sono indici che descrivono rispettivamente la posizione e la dispersione di una variabile casuale.

Osservazione 8.1: vogliamo qui provare una relazione approssimata di grande importanza pratica, che ha per la varianza lo stesso ruolo di quanto provato per la media nel Corollario 7.2 di questo quaderno.

Cominciamo ad osservare che se $Y = g(X)$ è una funzione della v.c. X, di distribuzione nota, la varianza della Y può essere calcolata, in base al teorema della media, come

$$\sigma^2(Y) = E\{(g(X) - \mu_Y)^2\}$$

intendendo $E\{\bullet\}$ come media su (X), ovvero come

$$\sigma^2(Y) = E\{Y^2\} - \mu_Y^2 = E\{g^2(X)\} - E^2\{g(X)\} \ . \tag{8.9}$$

Naturalmente la (8.9), che è una formula esatta, non è sempre di facile uso nei casi pratici; si è perciò pensato di cercare una formula approssimata di più semplice applicazione. Ci poniamo, a questo scopo, nelle stesse condizioni del Corollario 7.2: notiamo che, ora, per verificare che la variabile X sia concentrata basta supporre che σ_X sia piccolo, scegliendo poi ad esempio l'intervallo

$$[a,b] = [\mu_X - \lambda\sigma_X, \mu_X + \lambda\sigma_X]$$

con λ opportunamente grande.

Si porrà allora

$$\sigma^2(Y) \cong \int_a^b (g(x) - \mu_Y)^2 f_X(x) dx \ ;$$

ma in $[a, b]$

$$g(x) \cong g(\mu_X) + g'(\mu_X)(x - \mu_X) \cong \mu_Y + g'(\mu_X)(x - \mu_X)$$

così che

$$\begin{aligned} \sigma^2(Y) &\cong \int_a^b [g'(\mu_X)(x - \mu_X)]^2 f_X(x) dx \cong \\ &\cong [g'(\mu_X)]^2 \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx \quad . \end{aligned} \tag{8.10}$$

Ricordando la definizione di σ_X^2, la (8.10) ci dice che

$$\sigma_Y^2 \cong [g'(\mu_X)]^2 \ \sigma_X^2 \quad . \tag{8.11}$$

In pratica le (7.9) e (8.2) ci dicono che sotto opportune condizioni di concentrazione della variabile X e di regolarità della funzione $y = g(x)$, conoscendo μ_X, σ_X^2 è possibile calcolare direttamente, seppure in forma approssimata, μ_Y, σ_Y^2.

La (8.11) è conosciuta come *legge di propagazione della varianza.*

Osservazione 8.2: la (8.11), che in generale è una formula approssimata, diventa esatta quando $g(x)$ sia una funzione lineare $y = ax + b$. In tal caso infatti si ha (per la linearità della media)

$$\mu_Y = a\mu_X + b$$

e quindi

$$\begin{aligned} \sigma_Y^2 &= E\{(Y - \mu_Y)^2\} = E\{(aX + b - a\mu_X - b)^2\} = \\ &= a^2 E\{(X - \mu_X)^2\} = a^2 \sigma_Y^2 \ , \end{aligned} \tag{8.12}$$

c.v.d.

Osservazione 8.3: data una v.c. X qualsiasi è possibile, mediante una trasformazione lineare, costruire una variabile U che abbia media nulla e varianza $\sigma_U^2 = 1$.

Basta infatti porre

$$U = \frac{X - \mu_X}{\sigma_X} \ , \tag{8.13}$$

che si avrà (ricordando anche la (8.12))

$$E\{U\} = \frac{1}{\sigma_X} E\{(X - \mu_X)\} = 0$$
$$\sigma^2(U) = \frac{1}{\sigma_X^2} \cdot \sigma_X^2 = 1 \ .$$

La trasformazione (8.13) viene detta standardizzazione della v.c. e la v.c. U viene detta standardizzata.

9 Momenti - Funzione generatrice di momenti - Funzione caratteristica

Oltre a media e varianza vi sono altri indici per descrivere in modo sintetico la posizione e la forma della distribuzione di una v.c. In questo paragrafo ne prendiamo in considerazione alcuni.

Mediana. E' un indice di posizione ed è definito come quel valore c per cui (fig. 9.1a)

$$P(X \leq c) = P(X \geq c) = \frac{1}{2} \ ; \tag{9.1}$$

E' chiaro che se la distribuzione ammette un polo c di simmetria, allora c oltre ad essere la media è anche la mediana.

Oltre alla mediana vengono talvolta anche usati i quantili q_α definiti tramite la relazione

$$P(X \leq q_\alpha) = \alpha.$$

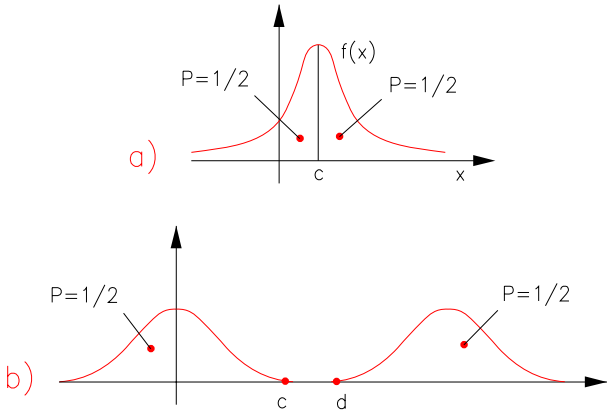


Figura 9.1

Si noti che la mediana non è necessariamente definita in modo univoco per ogni distribuzione; ad esempio nel caso rappresentato in fig. 9.1 b) tutti i valori tra c e d sono mediane.

Moda. E' un indice di posizione ed è definito come il valore c_0 in cui $f(x)$ raggiunge il suo valore massimo (assoluto). Nel caso il massimo assoluto sia raggiunto in più punti, la distribuzione è detta plurimodale e naturalmente la moda non è più univocamente definita: in caso contrario la distribuzione è detta unimodale (fig. 9.2).[9]

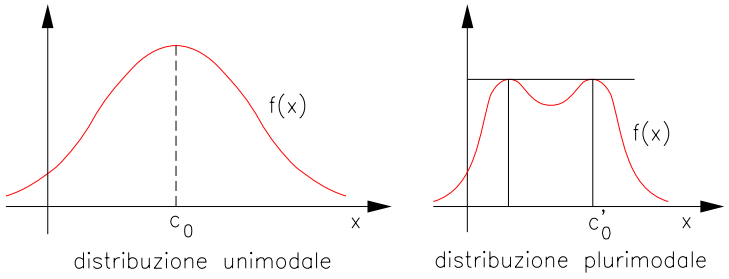


Figura 9.2

[9]Si può notare che di solito un modello probabilistico di tipo plurimodale sta a indicare un meccanismo stocastico composito in cui di volta in volta le cause seguono leggi diverse.

Momenti (semplici). E' un insieme di indici μ_n detti momenti di ordine n e definiti dalle relazioni

$$\mu_n = E\{X^n\} \tag{9.2}$$

Si può dimostrare che, sotto opportune condizioni, la conoscenza di tutti i momenti semplici $\{\mu_n\}$ è equivalente alla conoscenza di $f(x)$. Ciò ad esempio avviene se si sa a priori che vale una diseguaglianza del tipo $f(x) \leq Ce^{-\alpha x^2}$ per un qualche C e un qualche α positivi.

Si noti che μ_1 coincide con la media.

Momenti (centrali). Sono indici di forma che indicheremo col simbolo $\overline{\mu}_n$, definiti dalle relazioni

$$\overline{\mu}_n = E\{(X-\mu)^n\} \tag{9.3}$$

E' chiaro che

$$\begin{aligned} \overline{\mu}_1 &= E\{(X-\mu)\} = 0 \\ \overline{\mu}_2 &= E\{(X-\mu)^2\} = \sigma^2 \ . \end{aligned}$$

Tra momenti centrali di ordine n e momenti semplici sussiste la relazione

$$\begin{aligned} \overline{\mu}_n &= E\{(X-\mu)^n\} = E\{\sum_{k=0}^{n} \binom{n}{k} x^{n-k}(-\mu)^k\} = \\ &= \sum_{k=0}^{n} \binom{n}{k} \mu_{n-k}(-\mu)^k \ ; \end{aligned} \tag{9.4}$$

per $n = 2$ la (9.4) non è altro che la (8.6).

Osservazione 9.1: (skewness): è chiaro che, se una distribuzione è simmetrica, rispetto alla media, allora tutti i momenti centrali di ordine dispari sono nulli:

$$\overline{\mu}_{2k+1} = 0 \ ;$$

si può provare che è vero anche viceversa, ad esempio quando valgono le condizioni di equivalenza tra la conoscenza di tutti i μ_k e la conoscenza di $f(x)$.

Per conseguenza, l'annullarsi o meno degli indici dispari può essere assunto come misura del grado di simmetria della distribuzione.

In pratica l'indice più usato è

$$\beta = \frac{\overline{\mu}_3}{\sigma^3} \ , \tag{9.5}$$

detto anche indice di asimmetria o *skewness*.

Il vantaggio di tale indice è di essere invariante (a meno del segno) per trasformazioni lineari della X, cioè se

$$Y = aX + b$$

allora

$$|\beta_Y| = |\beta_X| \ ,$$

il che significa che β non dipende né da traslazioni della variabile, né da variazioni delle unità di misura.

Naturalmente l'annullarsi di β non è condizione sufficiente per la simmetria di $f(x)$, mentre $\beta \neq 0$ denuncia sicuramente una asimmetria.

Osservazione 9.2 (curtosi): un altro indice derivato dai momenti centrali ed utilizzato nella descrizione qualitativa delle distribuzioni è il cosiddetto indice di curtosi

$$\gamma = \frac{\overline{\mu}_4}{\sigma^4} \ ; \tag{9.6}$$

come è facile verificare anche tale indice è indipendente da trasformazioni lineari della variabile.

Applicando le definizioni si trova che per la normale standardizzata Z definita in (5.10), come per tutte le normali che da questa derivano per trasformazioni lineari, si ha

$$\gamma = \frac{4!}{4.2!} = 3 \ .$$

Le distribuzioni per cui $\gamma < 3$ vengono dette platicurtiche ed hanno delle code che tendono a zero più rapidamente della normale; le distribuzioni con $\gamma > 3$ sono chiamte leptocurtiche ed hanno viceversa delle code che vanno a zero più lentamente della normale. A mo' di esempio si calcoli il coefficiente γ per una v.c. uniformemente distribuita, ad esempio sull'intervallo $[1, 1]$, mostrando che in tal caso $\gamma = 1,8 < 3$. Al contrario una distribuzione (di Laplace), $f(x) = 1/2e^{-|x|}$, ha una curtosi $\gamma = 6$.

La funzione generatrice dei momenti. E' per definizione quando esiste

$$G_X(t) = E\{e^{tX}\} \ ; \tag{9.7}$$

L'indice X serve qui a indicare che si tratta della funzione generatrice dei momenti della v.c. X. Naturalmente si ha sempre

$$G_X(0) = 1 \ ;$$

é facile inoltre vedere che, se esiste per valori $t \neq 0$, $G_X(t)$ è definita in tutto un intervallo contenente $t = 0$, ed in tale intervallo è analitica (cioè sviluppabile in serie di Taylor).

La relazione tra $G_X(t)$ ed i momenti μ_n è facile da trovarsi notando che

$$\frac{d^n}{dt^n} G_X(t) = E\{X^n e^{tX}\}$$

così che

$$\mu_n = \frac{d^n}{dt^n} G_X(t)|_{t=0} \quad . \tag{9.8}$$

Viceversa, per l'analiticità di $G_X(t)$, si ha

$$G_X(t) = \sum_{n=0}^{+\infty} \frac{\mu_n}{n!} t^n \ , \tag{9.9}$$

per l'intervallo di convergenza della serie a secondo membro.

Vediamo ora come si trasforma la $G_X(t)$ per una trasformazione lineare di X,

$$Y = aX + b \ .$$

Si ha

$$\begin{aligned} G_Y(t) &= E\{e^{tY}\} = E\{e^{taX+tb}\} = e^{tb}E\{e^{atX}\} = \\ &= e^{bt}G_X(at) \ , \end{aligned} \tag{9.10}$$

che risolve il problema posto. Si noti che in particolare con la trasformazione

$$Y = X - \mu_X$$

si ha che

$$\mu_n(Y) = E\{Y^n\} = E\{(X - \mu_X)^n\} = \overline{\mu}_n(X) \ \ .$$

Ne deriva che la funzione generatrice dei momenti di Y sarà data da

$$G_Y(t) = \sum_{n=0}^{+\infty} \frac{\overline{\mu}_n}{n!} t^n \ . \tag{9.11}$$

Le funzioni generatrici dei momenti sono tra l'altro assai utili per il calcolo di tutti i momenti di una distribuzione.

Esempio 9.1: si consideri la v.c. Z distribuita secondo la densità di probabilità

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{z^2/2} \ ; \tag{9.12}$$

vogliamo in primo luogo calcolare $\mu_n(Z)$. Si ha, per definizione,

$$G_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tz} e^{z^2/2} dz =$$
$$= \frac{e^{\frac{1}{2}t^2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{(z-t)^2/2} dz = e^{t^2/2} \ ,$$

così che

$$G_Z(t) = \sum_{n=0}^{+\infty} \frac{1}{2^n} = \frac{t^{2n}}{n!} = \sum_{n=0}^{+\infty} \frac{(2n)!}{2^n n!} \frac{t^{2n}}{(2n)!}$$

Confrontando con la (9.9) si vede che

$$\mu_{2n+1} = 0 \quad \text{(in particolare } \ \mu_Z = 0)$$

come era evidente data la simmetria di $f(z)$, ed inoltre

$$\mu_{2n} = \frac{(2n)!}{2^n n!} \ . \tag{9.13}$$

Dalla (9.13) in particolare si ricava che

$$\mu_2 = \overline{\mu}_2 = \sigma^2(z) = \frac{2!}{2.1!} = 1 \quad .$$

cioè Z ha media nulla e varianza unitaria, essendo perciò una variabile standardizzata; è per questo motivo che, come si è già detto in altri esempi, Z viene chiamata la variabile normale (o gaussiana) standardizzata.

Supponiamo ora di considerare una trasformazione lineare di Z

$$X = aZ + b \ ;$$

ci chiediamo quale è la media e quali i momenti centrali di X.

Tenuto conto che $\mu_Z = 0$, per la linearità della media si ha subito

$$\mu_X = a\mu_Z + b = b \quad .$$

Pertanto sarà

$$\overline{\mu}_n(X) = E\{(X-\mu_X)^n\} = E\{(X-b)^n\} = E\{a^n Z^n\} = a^n \mu_n(Z) \ ,$$

ovvero

$$\begin{cases} \overline{\mu}_{2n}(X) = a^{2n} \frac{(2n)!}{2^n n!} \\ \overline{\mu}_{2n+1}(X) = 0 \ . \end{cases} \tag{9.14}$$

Dalla (9.13) per $n = 1$ risulta

$$\sigma^2(X) = \overline{\mu}_2 = a^2 \ .$$

Tenendo conto di queste relazioni tra a, b e σ_X^2, μ_X e ricordando la (6.7) (Esempio 6.1) si ha che la densità di probabilità di X può essere scritta come

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \quad ; \tag{9.15}$$

é questa la celebre formula di una distribuzione normale (o gaussiana) con media μ_X e varianza σ_X^2.

Si noti che la (9.11) coincide appunto con la (9.14) quando si prenda $\mu = 0, \ \sigma^2 = 1$. Data l'importanza di questa distribuzione, importanza che sarà meglio illustrata in seguito, si sono tabulati i valori della $f_Z(z)$ e della funzione di distribuzione $F_Z(z)$. Le corrispondenti funzioni per una distribuzione normale di media e varianza qualunque si possono ottenere mediante le trasformazioni di variabili.

$$X = \mu_X + \sigma_X^Z$$
$$Z = \frac{X - \mu_X}{\sigma_X}$$

Esempio 9.2: usando i dati dell'esempio 5.3, cioè la serie di 150 valori estratti da una normale standardizzata, vogliamo calcolare la media e i momenti di $2°, 3°$ e $4°$ ordine prima per le tre serie di 50 valori poi per i

150 valori complessivamente, per confrontarli con i corrispondenti valori teorici.

Per calcolare varianza e momenti centrali del 3° e 4° ordine, conviene prima calcolare i momenti semplici

$$\mu_k = \frac{1}{N}\sum_i (x_i)^k$$

e poi usare le formule (derivate dalla (9.4))

$$\begin{aligned}
\sigma^2 &= \mu_2 - \mu^2 \\
\overline{\mu}_3 &= \mu_3 - 3\mu_2\mu + 2\mu^3 \\
\overline{\mu}_4 &= \mu_4 - 4\mu_3\mu + 6\mu_2\mu^2 - 3\mu^4
\end{aligned}$$

Otteniamo così le tabelle:

	μ	μ_2	μ_3	μ_4
Serie 0	0,074	0,937	0,369	2,520
Serie 1	0,085	1,049	0,695	3,769
Serie 2	-0,078	0,831	-0,188	1,812
Serie 0+1+2	0,027	0,939	0,292	2,700
	(media teorica $\mu = 0$)			

	σ^2	$\overline{\mu}_3$	$\overline{\mu}_4$
Teorici	1,000	0,000	3,000
Serie 0	0,932	0,162	2,441
Serie 1	1,042	0,429	3,578
Serie 2	0,825	0,006	1,783
Serie 0+1+2	0,938	0,216	2,673

La funzione caratteristica. Essa è definita come

$$\begin{aligned}
C_X(t) &= E\{e^{itX}\} = \\
&= E\{\cos tX\} + iE\{\sin tx\} \ . \qquad (9.16)
\end{aligned}$$

Al contrario della funzione generatrice dei momenti, $C_X(t)$ esiste qualunque sia la funzione di distribuzione $F_X(x)$.

Accenniamo solo alla dimostrazione peraltro ben nota dalla teoria della trasformata di Fourier. In effetti se teniamo conto che $dF_X(x)$ è sempre positiva, si vede che presi N_1, M_1 ed N_2, M_2 positivi, chiamando $N = max(N_1,N_2), M = max(M_1,M_2), N' = min(N_1,N_2), M' = min(M_1,M_2)$, si trova

$$\begin{aligned}
&\left| \int_{-N_1}^{M_1} e^{itx} dF_X(x) - \int_{-N_2}^{M_2} e^{itx} dF_X(x) \right| = \qquad (9.17)\\
&= \left| \int_{-N_1}^{-N_2} e^{itx} dF_X(x) + \int_{M_2}^{M_1} e^{itx} dF_X(x) \right| \leq \\
&\leq \int_{-N}^{-N'} |e^{itx}| dF_X(x) + \int_{M'}^{M} |e^{itx}| dF_X(x) = \\
&= \int_{-N}^{-N'} dF_X(x) + \\
&+ \int_{M'}^{M} dF_X(x) = \{F_X(-N') - F_X(-N)\} + \{F_X(M') - F_X(M)\}
\end{aligned}$$

il che mostra che il primo membro della (9.17) tende a zero quando $N_1, N_2, M_1, M_2 \to \infty$, perché in tal caso $F_X(N') \to 0, F_X(N) \to 0$, $F_X(M') \to 1, F_X(M) \to 1$. Ne segue, per il criterio di Cauchy sull'esistenza dei limiti, che esiste sempre

$$\lim_{N,M \to \infty} \int_{-N}^{M} e^{itx} dF_X(x) = E\{e^{itx}\} \ . \qquad (9.18)$$

Procedendo come per la $G_X(t)$, si vede che, se esistono i momenti fino all'ordine n, allora

$$\mu_n = \frac{1}{i^n} \frac{d^n}{dt^n} C_X(t)|_{t=0} \ ; \qquad (9.19)$$

in tal caso si dimostra che

$$C_X(t) = \sum_{k=0}^{n} \frac{i^k \mu_k}{k!} t^k + \varepsilon_n(t) \frac{i^n t^n}{n!} \qquad (9.20)$$

dove è

$$\lim_{t\to 0} \varepsilon_n(t) = 0$$

Dunque l'esistenza dei momenti fino al grado n è legata alla regolarità di $C_X(t)$, nel senso che se $|\mu_n| < +\infty$ allora $C_X(t)$ è non solo continua, cosa che vale sempre, ma è anche derivabile n volte con continuità.

Ci si potrebbe chiedere se per una v.c. X che ammetta tutti i momenti, la (9.20) possa essere riscritta sotto forma di serie.

A tale riguardo vale il seguente teorema:

> **Teorema 9.1:** se X ammette momenti assoluti $|\mu|_n = E\{|X|^n\}$ finiti per tutti gli ordini e se per una qualche costante K
>
> $$\frac{|\mu|_n^{1/n}}{n} \leq K \tag{9.21}$$
>
> allora vale la rappresentazione
>
> $$C_X(t) = \sum_{k=0}^{+\infty} \frac{i^k \mu_k}{k!} t^k \ , \tag{9.22}$$

tale serie essendo convergente nell'intervallo $|t| < (eK)^{-1}$; inoltre la $C_X(t)$ definita dalla serie (9.22) solo su $|t| < (eK)^{-1}$ può essere continuata analiticamente a tutto l'asse $-\infty < t < +\infty$.

Notiamo che in conseguenza di questo teorema si può affermare che, quando vale la condizione (9.21), la conoscenza di tutti i μ_n porta alla conoscenza della $C_X(t)$ per tutti i t.

Ora ci si può anche chiedere se conoscendo la funzione caratteristica $C_X(t)$ non sia possibile ritrovare la funzione di distribuzione $F_X(x)$.

Ciò è possibile utilizzando in modo appropriato il seguente teorema:

> **Teorema 9.2:** per tutte le coppie di punti a, b in cui $F_X(x)$ risulta continua, si ha
>
> $$F_X(b) - F_X(a) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-ita} - e^{-itb}}{it} C_X(t) dt \ , \tag{9.23}$$

inoltre se risulta

$$\int_{-\infty}^{+\infty} |C_X(t)|dt < +\infty \ , \tag{9.24}$$

la v.c. X è continua ed ammette come densità di probabilità

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} C_X(t)dt \ . \tag{9.25}$$

Osservazione 9.3: combinando i Teoremi 9.1 e 9.2 si ha una risposta possibile al cosiddetto problema dei momenti che può così essere riassunto: dati i momenti μ_n di X è possibile ricostruire la corrispondente funzione di distribuzione $F_X(x)$? La risposta è che se vale la (9.21) i μ_n determinano $C_X(t)$ da cui si può ricavare la $F_X(x)$ tramite la (9.23). Questa risposta é ancora insoddisfacente perché se conosciamo solo μ_n non conosciamo invece $|\mu|_n$.

Tuttavia è possibile fornire una condizione sufficiente basata solo sui μ_{2n}, che naturalmente coincidono coi corrispondenti momenti assoluti. Vale dunque il

Lemma 9.1: se vale la sola condizione

$$\frac{(\mu_{2n})^{1/2n}}{2n} \leq K \tag{9.26}$$

allora i momenti μ_n determinano univocamente la $F_X(x)$.

Occorre anche notare che si conosce l'esempio di due v.c. con diverse funzioni di distribuzione X, per cui non vale la (9.26), che hanno tutti i momenti identici, mostrando così che al di fuori delle condizioni date non è detto che il problema dei momenti ammetta soluzione unica.

10 Alcune importanti variabili casuali

In questo paragrafo si raccolgono e descrivono alcune importanti famiglie di variabili casuali che ricorreranno in tutto il testo. Quanto ai meccanismi stocastici che generano tali variabili, per alcune di esse saranno presentati a livello di esercizi, mentre per altre occorre rimandare tale comprensione allo studio delle variabili a più dimensioni.

a) La distribuzione binomiale

Si consideri un esperimento stocastico $\mathcal{E}$ e siano S i suoi possibili risultati: dividiamo S in due insiemi A e B con

$$A \cup B = S \ , \ A \cap B = \emptyset \ .$$

Sia inoltre

$$P(A) = p \ , P(B) = q \ , \ p + q = 1 \ .$$

Consideriamo ora n ripetizioni indipendenti dell'esperimento $\mathcal{E}$: indichiamo con K la variabile casuale (discreta) che descrive la probabilità che su n esperimenti $\mathcal{E}, k$ abbiano dato un risultato in A (successo) ed $n-k$ abbiano dato un risultato in B (insuccesso).

Per trovare la distribuzione di K possiamo usare ricorsivamente la formula

$$P(n,k) = pP(n-1,k-1) + qP(n-1,k) \ ,$$

che dice che la probabilità di k successi su n prove è uguale alla probabilità di $k-1$ successi su $n-1$ prove per la probabilità di un nuovo successo (nell'n-esima prova), più la probabilità di k successi in $n-1$ prove per la probabilità di un insuccesso (nell'n-esima prova).

Partendo da

$$P(1,0) = q \ , \ P(1,1) = p$$

si trova

$$P(n,k) = \binom{n}{k} p^k \ q^{n-k} \quad (k = 0, 1, 2 \ldots) \ , \tag{10.1}$$

da cui il nome di distribuzione binomiale.

La funzione generatrice dei momenti di K è

$$G_K(t) = \sum_{k=0}^{n} \binom{n}{k} q^{n-k} p^k e^{kt} = (q + pe^t)^n$$

da cui si ricava facilmente

$$\begin{cases} \mu_K = np \\ \sigma_K^2 = \mu_2(K) - \mu^2(K) = npq \end{cases} \tag{10.2}$$

b) La distribuzione poissoniana

Si consideri una successione di variabili binomiali di numerosità n crescente e si supponga che la probabilità di "successo" di un evento elementare p cambi di n in n in modo tale che

$$np = \lambda \ \text{(costante)} \ .$$

Avremo in questo modo un modello stocastico che descrive il numero di successi k, su una sequenza al limite infinita di ripetizioni, quando la probabilità di un singolo evento è assai piccola, infinitesima.

La distribuzione limite della variabile K è la cosiddetta poissoniana.

$$P(K = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \ldots) \tag{10.3}$$

Infatti partendo dalla funzione generatrice dei momenti della binomiale, riscritta come

$$G_K(t) = [1 + p(e^t - 1)]^n$$

e ponendo $p = \dfrac{\lambda}{n}$ si trova al limite

$$G_K(t) = \left[1 + \frac{\lambda(e^t - 1)}{n}\right]^n \overset{(n\to\infty)}{\longrightarrow} e^{\lambda(e^t - 1)} \ .$$

Poiché questa è la funzione generatrice della (10.3) ne risulta l'assunto.

Si noti che

$$\begin{cases} \mu_K = \lambda \\ \sigma_K^2 = \lambda \ . \end{cases} \tag{10.4}$$

c) La distribuzone normale (o gaussiana)

Questa distribuzione è data dalla formula

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < +\infty) \ . \tag{10.5}$$

Come è già stato provato μ e σ^2 sono rispettivamente media e varianza: la funzione generatrice dei momenti è data da (vedi (9.10) ed esempio 9.1)

$$G_X(t) = e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2} \tag{10.6}$$

ed i momenti centrali (cfr. (9.14)).

$$\begin{aligned} \overline{\mu}_{2n+1} &= 0 \\ \overline{\mu}_{2n} &= \frac{\sigma^{2n}(2n)!}{2^n n!} \end{aligned} \tag{10.7}$$

Dalla normale standardizzata $Z(\mu = 0, \ \sigma = 1)$ si trovano nelle tavole i valori di $f_Z(z)$ e di

$$F(z) - \frac{1}{2} = \int_0^z f_Z(z) \ dz \ .$$

Una distribuzione normale, con opportuno μ e σ^2, è di solito un modello molto buono per rappresentare i processi di misura di alta precisione, in cui generalmente nessuna causa d'errore ha effetti significativamente più alti e riconoscibili delle altre concause.

d) La distribuzione χ^2

E' descritta dalla densità di probabilità, dipendente da un parametro

$$f(\chi^2) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{\nu}{2}-1} \quad (\chi^2 \geq 0) \ :^{10} \tag{10.8}$$

la (10.6) è detta distribuzione di una variabile χ^2 a ν gradi di libertà. La funzione generatrice dei momenti della χ^2_ν, è data da

[10] La funzione $\Gamma(p)$(Γ di Eulero) che compare nella (10.8) può essere definita per

$$G_{\chi^2_\nu}(t) = \frac{1}{(1-2t)^{\nu/2}} \quad : \tag{10.9}$$

dalla (10.8) si ricava facilmente

$$\mu(\chi^2_\nu) = \nu \tag{10.10}$$
$$\sigma^2(\chi^2_\nu) = 2\nu \ . \tag{10.11}$$

Nelle tavole sono riportati i valori $\chi^2_{(\alpha)}$ (per ν da 1 a 100) per cui

$$P(\chi^2 \leq \chi^2_{(\alpha)}) = \alpha \ ;$$

per $\nu > 100$ si può usare la formula approssimata

$$Z = \sqrt{2\chi^2} - \sqrt{2\nu - 1}$$

e per Z le tavole della normale standardizzata.

e) La distribuzione Γ

E' una variabile con densità di probabilità, dipendente dai parametri ρ e k,

$$f_\Gamma(x) = \frac{1}{\Gamma(k)} \rho^k x^{k-1} e^{-\rho x} \ , \ (x \geq 0) \ . \tag{10.12}$$

E' facile vedere che, quando in particolare k è un semintero, $k = \nu/2$, posto

$$2\rho\Gamma = \chi^2 \ ,$$

si trova una χ^2 a ν gradi di libertà.

---

valori interi e seminteri $1, 3/2, 2, 5/2$) tramite le relazioni

$$\begin{cases} \Gamma(p+1) = p\Gamma(p) \\ \Gamma(1) = 1 \\ \Gamma(3/2) = \sqrt{\pi/2} \end{cases} .$$

La funzione generatrice dei momenti di Γ è

$$G_\Gamma(t) = \frac{\rho^k}{(\rho - t)^k}$$

da cui si ricavano media e varianza

$$\begin{cases} \mu_\Gamma = \frac{k}{\rho} \\ \sigma^2(\Gamma) = \frac{k}{\rho^2} \end{cases} . \tag{10.13}$$

f) La distribuzione t (di Student)

Anche in questo caso si ha una distribuzione dipendente da un parametro ν, detto "gradi di libertà" della t

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})} \frac{1}{(1 + \frac{t^2}{\nu})^{\frac{\nu+1}{2}}} \tag{10.14}$$

Per simmetria è evidente che

$$\mu(t) = 0 \ , \tag{10.15}$$

mentre si può provare che

$$\sigma^2(t) = \frac{\nu}{\nu - 2} \qquad (\nu > 2) \ . \tag{10.16}$$

Nelle tavole si trovano (per ν da 1 a 120) i valori $t_{(\alpha)}$ per cui

$$P\left(t \leq t_{(\alpha)}\right) = 1 - \alpha \ .$$

Per grandi valori di ν si può porre approssimativamente

$$t = Z$$

essendo Z una variabile standardizzata.

g) La distribuzione F (di Fischer)

Si tratta di una variabile dipendente da due parametri, λ, ν detti gradi di libertà

$$\begin{aligned} f(F_{\lambda,\nu}) &= A_{\lambda,\nu}\frac{F^{\frac{\lambda-2}{2}}}{(\lambda F+\nu)^{\frac{\lambda+\nu}{2}}} \qquad (F \geq 0) \qquad (10.17) \\ A_{\lambda,\nu} &= \frac{\lambda^{1/2}\nu^{1/2}\Gamma(\frac{\lambda+\nu}{2})}{\Gamma(\frac{\lambda}{2})\Gamma(\frac{\nu}{2})} \ ; \end{aligned}$$

media e varianza di $F_{\lambda,\nu}$ sono date da

$$\mu_F = \frac{\nu}{\nu-2} \quad (\nu > 2) \qquad (10.18)$$

$$\sigma^2(F) = \frac{2\nu^2(\lambda+\nu-2)}{\lambda(\nu-2)^2(\nu-4)} \quad (\nu > 4) \ . \qquad (10.19)$$

I valori più comunemente tabulati sono $F_{0,95}$ ed $F_{0,99}$ per diversi valori di λ e ν.

La densità di probabilità $f_{\lambda,\nu}(F)$ verifica la proprietà

$$f_{\lambda,\nu}\left(\frac{1}{F}\right) = f_{\lambda,\nu}(F)F^2 \ .$$

Infatti

$$f_{\lambda,\nu}(F) = A_{\lambda,\nu}\frac{F^{\frac{\lambda}{2}-1}}{(\lambda F+\nu)^{(\lambda+\nu)/2}} \quad , \quad A_{\lambda,\nu} = A_{\nu,\lambda}$$

da cui segue

$$\begin{aligned} f_{\lambda,\nu}\left(\frac{1}{F}\right) &= A_{\lambda,\nu}\frac{(\frac{1}{F})^{\nu/2-1}}{(\nu\frac{1}{F}+\lambda)^{(\nu+\lambda)/2}} = A_{\lambda,\nu}\frac{(\frac{1}{F})^{\nu/2-1}}{(\frac{1}{F})^{(\nu+\lambda)/2}(\nu+\lambda F)^{(\nu+\lambda)/2}} \\ &= A_{\lambda,\nu}\frac{(\frac{1}{F})^{-\nu/2-1}}{(\nu+\lambda F)^{(\nu+\lambda)/2}} = A_{\lambda,\nu}\frac{F^{\nu/2-1}}{(\lambda F+\nu)^{(\nu+\lambda)/2}}F^2 \end{aligned}$$

Ricordiamo la (6.4) (per funzione iniettiva): $f_Y(y(x)) = f_X(x)/|y'(x)|$; con $y(x) = 1/x$ si ha $f_Y(1/x) = x^2 f_X(x)$.

Ne segue che, se la variabile $F_{\lambda,\nu}$ ha densità $f_{\lambda,\nu}$, la variabile $G = 1/F_{\lambda,\nu}$ ha densità $F^2 f_{\lambda,\nu} = f_{\nu,\lambda}$. Questo torna con il fatto che una variabile con densità $f_{\lambda,\nu}$ è, come vedremo, il rapporto fra una variabile con densità χ^2_λ/λ e una con densità χ^2_ν/ν (vedi §13).

$$P\{F \geq F_0\} = \int_{F_0}^{\infty} f_{\lambda,\nu}(F)dF = \int_0^{1/F_0} f_{\nu,\lambda}(G)dG = P\{0 \leq G \leq \frac{1}{F_0}\} \ . \tag{10.20}$$

11 La variabile casuale a n dimensioni

Incominciamo con questo paragrafo a generalizzare quanto visto sinora in una dimensione alle variabili n-dimensionali. Il punto di partenza sarà una variabile aleatoria descritta tramite una distribuzione di probabilità a n dimensioni. Ciò significa che ogni valore argomentale di tale variabile potrà essere indicato come un vettore a n componenti

$$\underline{x} = \begin{vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{vmatrix} \tag{11.1}$$

ovvero come un punto in uno spazio euclideo R_u con n assi cartesiani. L'insieme dei valori argomentali S sarà dunque un insieme in R_u in cui é definita la nostra distribuzione di probabilità.

Come di consueto indicheremo la variabile stessa con una lettera maiuscola $\underline{X}$; la sottolineatura indica che si tratta di una variabile vettoriale le cui componenti saranno altrettante v.c.

$$\underline{X} = \begin{vmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{vmatrix} \ .$$

Se la distribuzione di probabilità è concentrata esclusivamente su punti, cioè se esiste una successione di punti $\underline{x}_i$ tali che

$$\Sigma_i P(\underline{X} = \underline{x}_i) = 1 \ ,$$

allora la variabile si dice discreta, in caso contrario si dice continua.

Una variabile discreta sarà sempre rappresentabile per mezzo di tabelle di valori con n indici; ad esempio per una variabile doppia ($n = 2$) si potrà porre

	x_{11}	x_{12}	$\dots$	x_{1k}
x_{21}	p_{11}	p_{21}	$\dots$	p_{k1}
x_{22}	p_{12}	p_{22}	$\dots$	p_{k2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_{2h}	p_{1h}	p_{2h}	$\dots$	p_{kh}

$$p_{ij} = P(x_i = x_{1i}, x_2 = x_{2j}) \ .$$

Naturalmente una variabile discreta finita è sempre assimilabile formalmente ad una variabile statistica: nel caso a 2 dimensioni la corrispondenza si avrà con una v.s. doppia in cui alle probabilità p_{ij} andranno sostituite le frequenze relative

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

Le distribuzioni continue si possono dividere in singolari e regolari, a seconda che vi siano sottoinsiemi di S a misura euclidea nulla cui viene assegnata una probabilità positiva oppure no: indicando con $V(A)$ la misura euclidea di un insieme A, si hanno le seguenti situazioni

$$\begin{array}{llll} \exists A & , \ V(A) = 0 & , \ P(A) > 0 & : \text{distribuzione singolare} \\ \forall A & , \ P(A) = 0 & \Rightarrow \ V(A) = 0 & : \text{distribuzione regolare} \end{array}$$

A loro volta le distribuzioni singolari possono essere catalogate in base alle dimensioni del supporto: ad esempio per $n = 2$ una distribuzione

di probabilità concentrata su una retta $ax_1 + bx_2 = c$ sarà una distribuzione singolare con supporto a una dimensione; se invece la probabilità è concentrata in punti singoli si ha un supporto a dimensione zero e la variabile è puramente discreta.

Un esperimento aleatorio con la sua distribuzione di probabilità viene poi chiamato *variabile casuale* quando é definita la probabilità per ogni insieme del tipo $\{x_1 \leq x_{01}, \ldots x_n \leq x_{0n}\}$

$$\begin{aligned} P(X_1 \leq x_{01}, \ldots X_n \leq x_{0n}) &= F(x_{01}, x_{02}, \ldots x_{0n}) = \\ &= F(\underline{x}_0) \end{aligned} \tag{11.2}$$

La (11.2) definisce la *funzione di distribuzione* della v.c. $\underline{X}$.

In analogia a quanto visto nel caso monodimensionale si può definire una *funzione densità di probabilità* di $\underline{X}$, quando essa esiste, tramite la relazione

$$f(\underline{x}) = \lim_{\rho \to 0} \frac{P(A)}{V(A)} \tag{11.3}$$

dove $V(A)$ è la misura euclidea di A, ρ è il diametro dell'insieme A che tende a zero attorno al punto $\underline{x}$. La (11.3) può anche essere scritta

$$f(\underline{x}) = \frac{dP(\underline{x})}{dV(\underline{x})} \ , \tag{11.4}$$

essendo $dV(\underline{x})$ un elemento di volume di R_n attorno ad x: dalla definizione appare chiaro che per ogni insieme A di misura finita si avrà (per distribuzioni continue regolari)

$$P(\underline{X} \in A) = \int_A f(\underline{x}) dV(\underline{x}) \ . \tag{11.5}$$

In base alla (11.5) si avrà in particolare

$$F(x_{01}, \ldots x_{0n}) = \int_{-\infty}^{x_{01}} dx_1 \ldots \int_{-\infty}^{x_{0n}} dx_n f(\underline{x}) \ . \tag{11.6}$$

La (11.6) può essere invertita e si ha

$$f(\underline{x}) = \frac{\partial^n F(\underline{x})}{\partial x_1 \dots \partial x_n} \ . \tag{11.7}$$

Esempio 11.1: in un urna sono contenute 2 palline bianche (b, B) e due palline nere (n, N). Si vogliono scrivere le distribuzioni delle variabili casuali (discrete) che descrivono: a) l'estrazione in blocco di due palline; b) due estrazioni successive con sostituzione nell'ipotesi che i possibili risultati delle estrazioni siano equiprobabili.

Gli spazi dei valori argomentali dei due casi saranno dati da

a)

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

(senza sostituzione)

b)

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

(con sostituzione)

poiché nel primo caso si hanno 12 possibili risultati e nel secondo 16, si ha

a)

I estr. / II estr.	b	B	n	N
b	0	1/12	1/12	1/12
B	1/12	0	1/12	1/12
n	1/12	1/12	0	1/12
N	1/12	1/12	1/12	0

b)

I estr. / II estr.	b	B	n	N
b	1/16	1/16	1/16	1/16
B	1/16	1/16	1/16	1/16
n	1/16	1/16	1/16	1/16
N	1/16	1/16	1/16	1/16

Esempio 11.2: osservando un gran numero di tiri al bersaglio si sono potute raggiungere due conclusioni: a) in ogni anello del tiro a segno i colpi tendono a distribuirsi uniformememente; b) contando i punteggi effettuati si è visto che, indicando con r la distanza dal centro e con

$dr(r_0)$ il raggio di una corona di spessore infinitesimo e di raggio medio r_0,

$$P(r \in dr(r_0)) = \frac{r_0}{\sigma^2} e^{-\frac{r_0^2}{2\sigma^2}} dr \ , \tag{11.8}$$

dove σ^2 è un parametro caratteristico della bravura del tiratore. Si vuole trovare la distribuzione bidimensionale dei tiri.

Notiamo che la (11.8) in pratica ci fornisce la probabilità che il punto $\underline{x} = (\xi, \eta)$ si trovi nella corona circolare dC (fig. 11.1). Poiché nella corona circolare la probabilità è uniformemente distribuita, si avrà

$$\begin{aligned} P((\xi,\eta) \in d\omega) &= P(X \in dC)\frac{d\omega}{dC} = \\ &= P(X \in dC)\frac{d\theta}{2\pi} = \\ &= \frac{1}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} r dr \frac{d\theta}{2\pi} \ . \end{aligned}$$

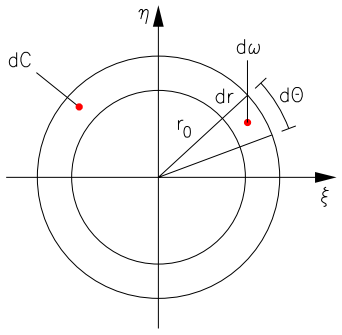


Figura 11.1

D'altro canto, per definizione di funzione di densità di probabilità

$$f(\xi,\eta) = \frac{P((\xi,\eta) \in d\omega)}{d\omega} = \frac{P((\xi,\eta) \in d\omega)}{r\ dr\ d\theta} = \frac{1}{2\pi\ \sigma^2} e^{-\frac{r^2}{2\sigma^2}} =$$

$$= \frac{1}{2\pi\sigma^2} e^{-\frac{\xi^2+\eta^2}{2\sigma^2}} \quad :$$

é questa, come vedremo tra breve, una distribuzione normale doppia.

12 Distribuzioni marginali - Distribuzioni condizionate - Dipendenza e indipendenza stocastica

Consideriamo l'evento [11]

$$A = \{X_1 \in dx_1(x_{01}),\ -\infty < X_2 < +\infty \ldots - \infty < X_n < +\infty\} \ ;$$

la classe di tali eventi dipende sostanzialmente da una sola variabile, x_{01}.

Pertanto, rispondendo alla domanda: quale è la probabilità che X_1 stia in $dx_1(x_{01})$ e $X_2, \ldots, X_n$ abbiano un valore qualunque, si genera una distribuzione unidimensionale ed una corrispondente v.c. X_1 definita dalla relazione

$$\begin{aligned} P(X_1 \in dx_1) &= P(\underline{X} \in A) = \\ &= dx_1 \int_{-\infty}^{+\infty} dx_2 \ldots \int_{-\infty}^{+\infty} dx_n f_X(x_{01}, x_2 \ldots x_n) \ . \end{aligned}$$

Questa v.c. è detta marginale della X ed ha densità di probabilità

$$\begin{aligned} f_{X_1}(x_{01}) = \frac{P(\underline{X} \in A)}{dx_1} &= \int_{-\infty}^{+\infty} dx_2 \ldots \int_{-\infty}^{+\infty} dx_n f_X(x_{01}, x_2 \ldots x_n) = \\ &= \frac{\partial}{\partial x_1} F_X(x_{01}, +\infty, \ldots + \infty) \ . \end{aligned} \tag{12.1}$$

Naturalmente una variabile n-dimensionale avrà n distribuzioni marginali. Oltre alle distribuzioni marginali di una sola componente per volta, si possono anche introdurre le distribuzioni marginali per coppie di

[11] Quando necessario useremo la scrittura $dx(P)$ per indicare che si tratta del differenziale x nel punto P; così $dx_1(x_{01})$ è il differenziale di x_1 nel valore x_{01} ovvero un intervallino infinitesimo dell'asse x_1 attorno ad x_{01}.

componenti $(x_1, x_2), (x_1, x_3) \ldots$ o per triplette di componenti ecc.: ad esempio

$$f_{X_1 X_2}(x_1, x_2) = \int_{-\infty}^{+\infty} dx_3 \ldots \int_{-\infty}^{+\infty} dx_n f_X(x_1, x_2, x_3 \ldots x_n) \; . \qquad (12.2)$$

In sostanza le distribuzioni marginali rispondono alla domanda: quale è la probabilità che un certo gruppo di componenti $x_{i_1}, \ldots x_{i_m}$ appartengano a un certo elemento di volume dV_m, indipendentemente dal valore assunto dalle altre componenti?

Oltre a questo, un altro problema è di una certa rilevanza e dà luogo a importanti distribuzioni: il problema è di conoscere la probabilità che m variabili (diciamo $x_1 \ldots x_m$) stiano in un elemento di volume dV_m; mentre le altre $(x_{m+1} \ldots x_n)$ sono certamente vincolate ad un elemento di volume dV_{n-m}.

In pratica, definendo i due eventi (fig. 12.1)

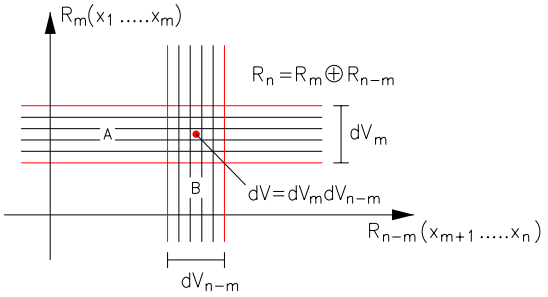


Figura 12.1

$$A = \{(X_1, \ldots X_m) \in dV_m\} \; , \; B = \{(X_{m+1}, \ldots X_n) \in dV_{n-m}\}$$

si vuole calcolare

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

$$= \frac{f_X(x_1 \dots x_m, x_{m+1} \dots x_n) dx_1 \dots dx_m}{\int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_m f_X(x_1 \dots x_m, x_{m+1} \dots x_n)} \tag{12.3}$$

Questa distribuzione di probabilità dà luogo ad una densità di probabilità per le $(x_1 \dots x_m)$, per ogni valore assegnato a $(x_{m+1} \dots x_n)$: questa distribuzione si chiama appunto distribuzione delle $(x_1 \dots x_m)$ condizionate a $(x_{m+1} \dots x_n)$ ed è data da

$$f_{X_1 \dots X_m | X_{m+1} \dots X_n}(x_1 \dots x_m | x_{m+1} \dots x_n) = \tag{12.4}$$

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

$$= \frac{f_X(x_1 \dots x_m, x_{m+1} \dots x_n)}{f_{X_{m+1} \dots X_n}(x_{m+1} \dots x_n)} \ .$$

In particolare la distribuzione di una componente, ad esempio X_1, condizionata alle altre, $(x_2 \dots x_n)$ sarà data da

$$f_{X_1 | X_2 \dots X_n}(x_1 | x_2 \dots x_n) = \frac{f_X(\underline{x})}{f_{X_2 \dots X_n}(x_2 \dots x_n)} \tag{12.5}$$

Vogliamo ora applicare il concetto di indipendenza stocastica ai due eventi A e B; per definizione A e B sono *stocasticamente indipendenti* se

$$P(A|B) = P(A)$$

Tenendo conto che

$$P(A) = P\left((X_1 \dots X_m) \in dV_m\right) = f_{X_1 \dots X_m}(x_1 \dots x_m) dV_m$$

si trova

$$\begin{aligned} & f_{X_1 \dots X_m | X_{m+1} \dots X_n}(x_1 \dots x_m | x_{m+1} \dots x_n) = \\ & = \frac{f_X(\underline{x})}{f_{X_{m+1} \dots X_n}(x_{m+1} \dots x_n)} = \\ & = f_{X_1 \dots X_m}(x_1 \dots x_m) \ : \end{aligned} \tag{12.6}$$

quando la (12.6) è verificata si dice che le $(x_1 \dots x_m)$ sono *stocasticamente indipendenti* dalle $(x_{m+1} \dots x_n)$.

Viceversa se la densità di probabilità globale $f_X(\underline{x})$ può essere fattorizzata in un prodotto del tipo

$$f_X(\underline{x}) = \varphi(x_1 \dots x_m)\psi(x_{m+1} \dots x_n) \tag{12.7}$$

si vede, in primo luogo, che φ e ψ sono proporzionali alle marginali rispettive.

Infatti, integrando la (12.7) si ottiene

$$\int_{-\infty}^{+\infty} dx_1 \dots dx_m f_X(\underline{x}) = f_{X_{m+1} \dots X_n}(x_{m+1} \dots x_n) =$$
$$= \left[\int_{-\infty}^{+\infty} dV_m \varphi(x_1 \dots x_m) \right] \cdot \psi(x_{m+1} \dots x_n) \ ,$$

cioè ψ è proporzionale alla $f_{X_{m+1} \dots X_n}$ e analogamente φ è proporzionale alla $f_{X_1 \dots X_m}$.

Pertanto, eventualmente riaggiustando le costanti di proporzionalità, dalla (12.7) discende

$$f_X(\underline{x}) = f_{X_1 \dots X_m}(x_1 \dots x_m) f_{X_{m+1} \dots X_n}(x_{m+1} \dots x_n) \ ,$$

così che la (12.6) risulta verificata.

Si giunge così al seguente teorema:

> **Teorema 12.1:** condizione necessaria e sufficiente affinché $(X_1 \dots X_m)$ siano stocasticamente indipendenti da $(X_{m+1} \dots X_n)$ è che la densità di probabilità congiunta si spacchi nel prodotto delle due marginali.

$$f_X(\underline{x}) = f_{X_1 \dots X_m}(x_1 \dots x_m) f_{X_{m+1} \dots X_n}(x_{m+1} \dots x_n) \ , \tag{12.8}$$

Da questo teorema, usando ripetutamente la (12.5) deriva facilmente un corollario.

Corollario 12.1: condizione necessaria e sufficiente affinché le componenti di una v.c. n-dimensionale siano tra loro stocasticamente indipendenti è che la densità di probabilità congiunta si spacchi nel prodotto delle n marginali

$$f_X(\underline{x}) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n) \qquad (12.9)$$

Esempio 12.1: consideriamo la distribuzione a due dimensioni

$$f(x,y) = \begin{cases} \dfrac{3}{2\pi R^3} \quad \sqrt{R^2 - x^2 - y^2} & (x^2 + y^2 \leq R^2) \\ 0 & (x^2 + y^2 > R^2) \ . \end{cases}$$

definita nel cerchio con centro nell'origine e di raggio R. Vogliamo calcolare le due distribuzioni marginali $f_X(x)$, $f_Y(y)$ e la distribuzione della Y condizionata ad X, $f(y|x)$.

Si ha (fig. 12.2):

$$f_X(x) = \int_{-\overline{y}}^{\overline{y}} dy \ f(x,y) = \frac{3}{2\pi R^3} \cdot 2 \int_0^{\overline{y}} \sqrt{R^2 - x^2 - y^2} \ dy \ ;$$

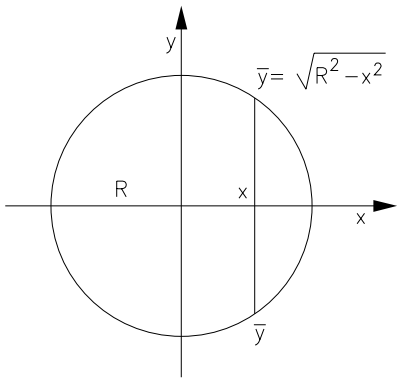


Figura 12.2

Con la sostituzione

$$y = \overline{y}\ \sin\ t \qquad (\overline{y} = \sqrt{R^2 - x^2}), 0 \leq t \leq \pi/2\ ,$$

si trova facilmente

$$f_X(x) = \frac{3}{2\pi R^3}\overline{y}^2 \cdot \frac{\pi}{2} = \frac{3}{4R^3}(R^2 - x^2) \qquad (-R \leq x \leq R)\ .$$

Per simmetria si ha anche

$$f_Y(y) = \frac{3}{4R^3}(R^2 - y^2) \qquad (-R \leq y \leq R)\ .$$

Si noti che, essendo

$$f(x, y) \neq f_X(x) f_Y(y)$$

le due variabili X e Y non sono tra loro stocasticamente indipendenti.

Quanto alla distribuzione della Y condizionata ad X si ha

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{2}{\pi}\frac{\sqrt{R^2 - x^2 - y^2}}{R^2 - x^2}\ .$$

Si può osservare che in sostanza la $f(y|\overline{x})$ ha la stessa forma, come funzione della y, della $f(y, \overline{x})$, corretta per un opportuno fattore che dipende solo da $\overline{x}$: nel nostro caso il grafico della $f(y|x)$ come funzione di y per $\overline{x}$ fissato, è una mezza ellisse (fig. 12.3).

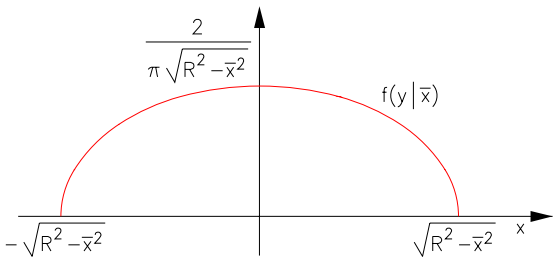


Figura 12.3

13 Trasformazioni di variabili

Supponiamo che sia data una trasformazione da R_n a R_m

$$\underline{y} = \underline{g}(\underline{x}) \tag{13.1}$$

ovvero:

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

A partire da una distribuzione di probabilità in R_n possiamo costruirne una in R_m secondo la seguente definizione: sia $dV_m(\underline{y}_0)$ un elemento di volume di R_m attorno ad $\underline{y}_0$ e sia $A_{\underline{y}_0}$ l'immagine inversa di $dV_m(\underline{y}_0)$ ovvero l'insieme di tutti quegli $\underline{x} \in R_n$ per cui si ha $\underline{g}(\underline{x}) \in dV_m(\underline{y}_0)$; poniamo allora:

$$P(\underline{Y} \in dV_m(\underline{y}_0)) = P(\underline{X} \in A_{\underline{y}_0}) \ . \tag{13.2}$$

Naturalmente ciò presuppone che la famiglia di insiemi di tipo $A_{\underline{y}_0}$ sia misurabile rispetto a P, cioè che si sappia dire quanto vale $P(A_{\underline{y}_0})$. In tal modo a partire da una v.c. $\underline{X}$ possiamo costruirne un'altra $\underline{Y}$ che indicheremo con la relazione

$$\underline{Y} = \underline{g}(\underline{X}) \quad .$$

Ci chiediamo ora come sia distribuito il vettore $\underline{Y}$, conoscendo la distribuzione di $\underline{X}$.

Si possono avere tre casi: $m = n$ e $\underline{g}(\underline{x})$ una trasformazione regolare, cioè con Jacobiano continuo e tale che

$$\det \left[\frac{\partial g}{\partial x}\right] \neq 0,$$

oppure $m < n$, o ancora $m > n$.

Quest'ultimo caso lo escludiamo, poiché se $\underline{g}(\underline{x})$ è differenziabile, per $m > n$ si ha che l'insieme dei valori argomentali $\underline{y} = \underline{g}(\underline{x}) \quad (\underline{x} \in R_n)$ costituisce un insieme a misura nulla in R_m e perciò si avrebbe una distribuzione singolare che noi qui non esamineremo.

a) $m = n$, $\underline{g}(\underline{x})$ **trasformazione regolare**

Per ipotesi $\underline{g}$ è differenziabile

$$\det \left[\frac{\partial g}{\partial x}\right] = \det \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix} \neq 0 \quad \forall \underline{x} \in R_n \ ,$$

ed esiste una corrispondenza biunivoca tra $\underline{y}$ e $\underline{x}$. Sia ora $dV_n(\underline{y})$ un elemento di volume attorno ad $\underline{y}$ e sia $dV_n(\underline{x})$ l'elemento di volume corrispondente, tale cioè che

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

per la definizione (13.2), si ha:

$$P\left(\underline{Y} \in dV_n(\underline{y})\right) = P(\underline{X} \in dV_n(\underline{x})) \ .$$

Per definizione di densità di probabilità è allora

$$f_Y(\underline{y}) dV_n(\underline{y}) \cong f_X(\underline{x}) dV_n(\underline{x})$$

ovvero:

$$f_Y(\underline{y}) = \frac{f_X(\underline{x})}{\dfrac{dV_n(\underline{y})}{dV_n(\underline{x})}}$$

Ricordando che per noti teoremi

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

si trova

$$f_Y(\underline{y}) = \frac{f_X(\underline{x})}{\left|\dfrac{\partial g}{\partial x}\right|} \quad (\text{dove } \underline{x} = \underline{g}^{(-1))}(\underline{y}) \tag{13.3}$$

La (13.3) è la naturale generalizzazione della (6.4).

Esempio 13.1: il caso più semplice di una trasformazione da R_n a R_n è naturalmente quello lineare

$$\underline{y} = A\underline{x} + \underline{b} \ , \tag{13.4}$$

dove per ipotesi

$$|A| = \ \det \ A = \left|\frac{\partial g}{\partial x}\right| \neq 0 \ , \tag{13.5}$$

così che esiste la trasformazione inversa

$$\underline{x} = A^{-1}(\underline{y} - \underline{b}) \tag{13.6}$$

Per la (13.3), se $f_X(\underline{x})$ è la distribuzione della X,

$$f_Y(\underline{y}) = \frac{f_X\left(A^{-1}(\underline{y} - \underline{b})\right)}{|A|} \tag{13.7}$$

Se ad esempio la $f_X(\underline{x})$ è il prodotto di n normali standardizzate, si potrà scrivere [12]

$$f_X(\underline{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \ldots \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}} =$$

[12] Ricordiamo che il trasposto di $\underline{x}$, $\underline{x}^+$ è un vettore riga $[x_1, \ldots, x_n]$ e che si ha

$$\underline{x}^+ \underline{x} = [x_1 \ldots x_n] \begin{vmatrix} x_1 \\ \vdots \\ x_n \end{vmatrix} = x_1^2 + \ldots + x_n^2 = |x|^2 \ .$$

$$f_X(\underline{x}) = \frac{1}{(2\pi)^{n/2}} e^{-\sum_1^n \frac{x_i^2}{2}} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{(\underline{x}^+ \underline{x})}{2}}$$

Notando che[13]

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

per la (13.7) si ha

$$\begin{aligned} f_Y(\underline{y}) &= \frac{1}{(2\pi)^{n/2}|A|} e^{-\frac{1}{2}(\underline{x}-\underline{b})^+ (A^+)^{-1} A^{-1} (\underline{x}-\underline{b})} = \\ &= \frac{1}{(2\pi)^{n/2}|A|} e^{-\frac{1}{2}(\underline{x}-\underline{b})^+ (AA^+)^{-1} (\underline{x}-\underline{b})} = \end{aligned} \tag{13.8}$$

Se poi A è una matrice simmetrica,

$$A^+ = A$$

la (13.18) diventa

$$f_Y(\underline{y}) = \frac{1}{(2\pi)^{n/2}|A|} e^{-\frac{1}{2}(\underline{x}-\underline{b})^+ (A^2)^{-1} (\underline{x}-\underline{b})} \ . \tag{13.9}$$

b) $m < n$

In questo caso in genere ad un elemento di volume $dV_m(\underline{y})$ può corrispondere un insieme di $\underline{x}, A_{\underline{y}}(dV_m)$, che non ha misura finita

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

Perciò usando la solita definizione di funzione di (13.2) si potrà tutt'al più scrivere

$$f_Y(\underline{y}) = \frac{1}{dV_m(\underline{y})} \int_{A_{\underline{y}}(dV_m)} f_X(\underline{x}) dV_n(\underline{x}) \tag{13.10}$$

[13]Ricordiamo che $(AB)^+ = B^+ A^+$, e che $(A^{-1})^+ = (A^+)^{-1}$.

La (13.10) andrà poi specificata caso per caso. Vedremo qui alcuni importanti casi.

b1) $m = 1, n = 2$

$$y = x_1 + x_2$$

In questo caso l'elemento di volume $dV_m(y)$ è dy mentre A_y sarà dato da (fig. 13.1)

$$y \leq x_1 + x_2 \leq y + dy \ .$$

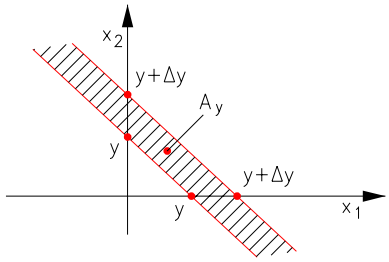


Figura 13.1

Notiamo che un elemento d'area di A_y può essere descritto come in fig. 13.2

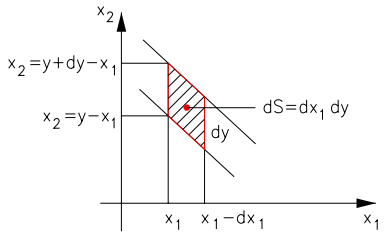


Figura 13.2

si ha

$$f_Y(y) = \frac{1}{dy}\int_{-\infty}^{+\infty} f_X(x_1, y - x_1)dx_1 dy = \tag{13.11}$$
$$= \int_{-\infty}^{+\infty} f_X(x_1, y - x_1)dx_1 \ .$$

Naturalmente si poteva usare x_2 come variabile corrente e si sarebbe ottenuto in questo caso

$$f_Y(y) = \int_{-\infty}^{+\infty} f_X(y - x_2, x_2)dx_2 \ . \tag{13.12}$$

La (13.11) assume poi la forma di integrale di convoluzione nel caso che X_1 e X_2 siano indipendenti

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) f_{X_2}(y - x_1)dx_1 \ . \tag{13.13}$$

La (13.13) è facilmente generalizzabile al caso n-dimensionale che, sempre nell'ipotesi di indipendenza di tutte le componenti, si presenta nella forma

$$\begin{aligned} &Y = X_1 + X_2 + \ldots + X_n \\ &f_X(\underline{x}) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n) \\ &f_Y(y) = \int_{-\infty}^{+\infty} dx_1 f_{X_1}(x_1) \int_{-\infty}^{+\infty} dx_2 f_{X_2}(x_2 - x_1) \ldots \\ &\ldots \int_{-\infty}^{+\infty} dx_{n-1} f_{X_{n-1}}(X_{n-1} - X_{n-2}) f_{X_n}(y - x_{n-1}) \ . \end{aligned} \tag{13.14}$$

b2) $m = 1, n = 2$

$$y = \frac{x_2}{x_1} \qquad (x_1 \geq 0)$$

Supponiamo dunque che

$$x = \left| \begin{array}{c} x_1 \\ x_2 \end{array} \right|$$

sia una v.c. distribuita sul semipiano $x_1 \geq 0$.

Poiché $m = 1, dV_m(y) = dy$ ed il corrispondente insieme $A_y(dy)$ è dato da (fig. 13.3)

$$x_1 y \leq x_2 \leq x_1(y + dy)$$

Poiché l'elemento d'area di A_y è dato da (fig. 13.3)

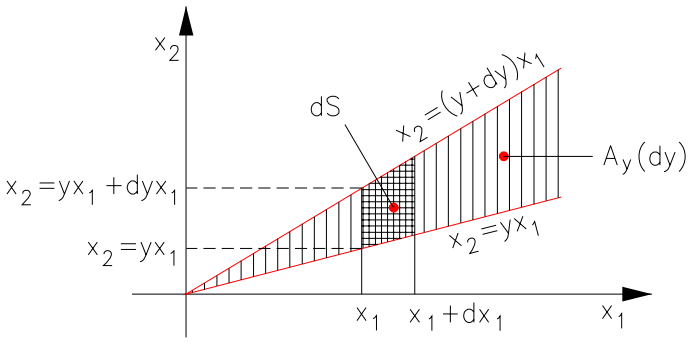


Figura 13.3

$$dS = dx_1 \cdot x_1 dy$$

si trova applicando la (13.10)

$$\begin{aligned} f_Y(y) &= \frac{1}{dy} \int_{A_y} f(x_1, x_2) dS = \\ &= \frac{1}{dy} \int_0^{+\infty} f(x_1, yx_1) dx_1 \cdot x_1 dy = \\ &= \int_0^{+\infty} f(x_1, yx_1) x_1 dx_1 \ . \end{aligned} \tag{13.15}$$

b3) $m = 1, n$ **qualsiasi**

$$y = \sqrt{x_1^2 + \ldots + x_n^2}$$

Supponiamo inoltre che la distribuzione congiunta delle $(x_1, \ldots, x_n)$ sia funzione solo di y, ovvero

$$f_X(x_1, \ldots, x_n) = f_X\left(\sqrt{x_1^2 + \ldots + x_n^2}\right) = f_X(y) \ .$$

Ora notiamo che l'insieme $A_y(dy)$ corrispondente ad un elemento dy non é altro che la corona sferica (di R_n) compresa tra $r = y$ ed $r = y + dy$ (fig. 13.4).

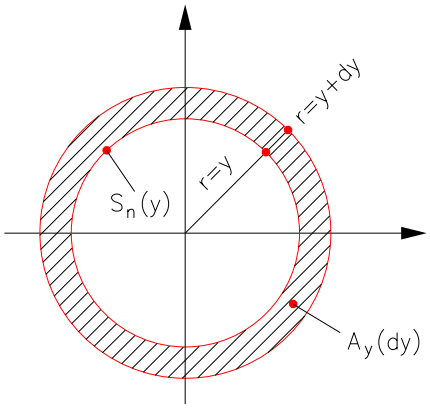


Figura 13.4

Poiché f_X è funzione di y essa sarà costante in A_Y e si avrà

$$\begin{aligned} f_Y(y) &= \frac{1}{dy}\int_{A_y} f_X(x_1, \ldots, x_n) dV_n = \\ &= \frac{1}{dy}\int_{A_y} f_X(y) dV_n = \\ &= f_X(y)\frac{d\omega(A_y)}{dy} \quad (y \geq 0) \end{aligned}$$

essendo $d\omega(A_y)$ la misura euclidea di $A_y(dy)$ può essere scritto

$$d\omega(A_y) = S_n(y)dy$$

essendo $S_n(y)$ la superficie della sfera di raggio y, così che

$$\frac{d\omega(A_y)}{dy} = S_n(y) \ .$$

Osserviamo infine che in R_n vale la relazione

$$S_n(y) = y^{n-1}S_n(1)$$

dove $S_n(1)$ (la superficie della sfera di R_n con raggio unitario) è una costante dipendente solo da n.

Pertanto riassumendo si ha

$$f_Y(y) = S_n(1)f_X(y)y^{n-1} \quad (y \geq 0) \ . \tag{13.16}$$

La costante $S_n(1)$ è data dalla formula

$$S_n(1) = 2\frac{\pi^{n/2}}{\Gamma(n/2)} \tag{13.17}$$

dove le quantità $\Gamma(n/2)$ sono già state definite (vedi nota a pag. 68).

Osservazione 13.1: spesso, in quanto segue, avremo bisogno della distribuzione di ξ

$$\xi = Y^2 = X_1^2 + \ldots + X_n^2 \ .$$

Usando la (6.8) si ha

$$\begin{aligned} f_\xi(\xi) &= \frac{f_Y(\sqrt{\xi}) + f_Y(-\sqrt{\xi})}{2\sqrt{\xi}} = \frac{f_Y(\sqrt{\xi})}{2\sqrt{\xi}} = \\ &= \frac{S_n(1)}{2} f_X(\sqrt{\xi})\xi^{(n/2)-1} \end{aligned} \tag{13.18}$$

Esempio 13.2: siano $z_1, \ldots, z_n,\ n$ variabili normali standardizzate indipendenti; vogliamo provare che

$$\xi = Z_1^2 + \ldots + Z_n^2$$

é una variabile χ^2 a n gradi di libertà.

Infatti, dalla (13.18), tenuto conto che

$$f_Z(\underline{z}) = \frac{1}{(2\pi)^{n/2}} e^{-(\sum_{i=1}^{n} z_i^2)/2} = \frac{1}{(2\pi)^{n/2}} e^{-\xi/2}$$

si trova

$$f_\xi(\xi) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-\xi/2} \xi^{(n/2)-1}$$

che appunto coincide con la (10.8).

Notiamo che come conseguenza di questo risultato si ha il notevole *teorema per la somma di variabili χ^2 indipendenti:*

$$\chi^2_{\nu_1} + \chi^2_{\nu_2} = \chi^2_{\nu_1+\nu_2}\ . \tag{13.19}$$

La (13.19) in sostanza dice che sommare due χ^2 indipendenti, rispettivamente a ν_1 e ν_2 gradi di libertà, è come sommare i quadrati di $\nu_1 + \nu_2$ normali standardizzate indipendenti.

Si può notare anche che in una certa misura il teorema può anche essere invertito: in effetti si può dimostrare che se ad una χ^2_ν si somma una variabile ξ non negativa, indipendente da questa e se il risultato è una $\chi^2_{\nu_1+\nu_2}$ allora ξ è necessariamente una $\xi^2_{\nu_2}$.

Esempio 13.3: sia Z una normale standardizzata e sia

$$Y = \sqrt{\chi^2_\nu} = \sqrt{Z_1^2 + \ldots + Z_\nu^2}\ ;$$

se Z ed Y sono tra loro indipendenti, allora la variabile

$$t = \frac{\sqrt{\nu} Z}{Y} \tag{13.20}$$

risulta essere una t di Student a ν gradi di libertà.

Infatti notiamo che per la (13.15) e per l'indipendenza di Z ed Y deve essere

$$f_t(t) = \int_0^{+\infty} dx \cdot x f_Y(x) f_{\sqrt{\nu} Z}(tx) \ .$$

Ora per la (13.16)

$$f_Y(x) = c_1 e^{-x^2/2} x^{\nu - 1}$$

inoltre $\sqrt{\nu} Z$ è una normale con media nulla e varianza ν, $N[0, \nu]$ così che

$$f_{\sqrt{\nu} Z}(tx) = \frac{1}{\sqrt{2\pi}\sqrt{\nu}} e^{-(tx)^2/2\nu} \ .$$

Perciò

$$f_t(t) = c_2 \int_0^{+\infty} dx \ x^{\nu} e^{-1/2(1+t^2/\nu)x^2}$$

con la sostituzione

$$\eta = \left(1 + \frac{t^2}{\nu}\right)^{1/2} x$$

si trova

$$\begin{aligned} f_t(t) &= \frac{c_2}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \int_0^{+\infty} d\eta \ \eta^{\nu} e^{-(1/2)\eta^2 =} \\ &= \frac{c_3}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \qquad (-\infty < t < +\infty) \end{aligned} \tag{13.21}$$

Poiché è noto a priori che

$$\int_{-\infty}^{+\infty} f_t(t)\ dt = 1$$

la (13.21) è univocamente determinata e non può che coincidere con la (10.14).

Esempio 13.4: siano date due variabili χ^2 a λ e ν gradi di libertà ed indipendenti tra loro: allora il rapporto

$$F = \frac{\frac{\chi^2_\lambda}{\lambda}}{\frac{\chi^2_\nu}{\nu}} \tag{13.22}$$

é una variabile F di Fischer a λ e ν gradi di libertà.

In effetti per la (13.15) sarà

$$f_F(F) = \int_0^{+\infty} dx\ x\ f_{X_1}(x) f_{X_2}(Fx)$$

dove

$$\begin{cases} X_1 = \chi^2_\nu/\nu \\ f_{X_1} = c_1 e^{-\nu x/2} x^{(\nu/2)-1} \end{cases}$$

$$\begin{cases} X_2 = \chi^2_\lambda/\lambda \\ f_{X_2} = c_2 e^{-\lambda x/2} x^{(\lambda/2)-1} \end{cases} .$$

Si ha quindi

$$\begin{aligned} f_F(F) &= c_1 c_2 \int_0^{+\infty} dx F^{\frac{\lambda}{2}-1} x^{\frac{\lambda+\nu}{2}-1} e^{(-\nu x)/2} e^{(-\lambda F x)/2} = \\ &= c_1 c_2 F^{\frac{\lambda}{2}-1} \int_0^{+\infty} dx\ x^{\frac{\lambda+\nu}{2}-1} e^{-1/2(\nu+\lambda F)x} \quad : \end{aligned}$$

con la sostituzione

$$\eta = (\nu + \lambda F)x$$

si trova quindi

$$f_F(F) = c_3 \frac{F^{\frac{\lambda}{2}-1}}{(\nu + \lambda F)^{(\lambda+\nu)/2}} \ ,$$

che, dovendo soddisfare alla condizione di normalizzazione

$$\int_0^{+\infty} f_F(F)\ dF = 1$$

coincide necessariamente con la (10.17).

Osservazione 13.2: si noti che scambiando λ con ν nella (13.22) si ottiene la quantità $1/F$: di conseguenza risulta evidente il risultato espresso da (10.20) .

14 Media - Covarianza

Anche per la variabile a n dimensioni si possono generalizzare i concetti già visti per una variabile a 1 dimensione.

Il primo concetto è quello di media della v.c. casuale X: per definizione la media di $\underline{X}$ è un vettore n dimensionale μ_X (se esiste), dato da

$$\mu_X = E\{\underline{X}\} = \int_{R_n} dV_n(\underline{x})\underline{x} f_X(\underline{x})$$

la componente i-esima di μ_X è data da

$$\mu_{X_i} = E\{x_i\} = \int_{-\infty}^{+\infty} dV_n(\underline{x}) x_i f_X(\underline{x}) \ : \tag{14.1}$$

Notiamo che per calcolare μ_{X_i} basta la distribuzione marginale di X_i, infatti

$$\begin{aligned}\mu_{X_i} &= \int_{R_n} dx_i \cdot dV_{n-1} x_i f_X(\underline{x}) = \\ &= \int_{-\infty}^{+\infty} dx_i \cdot x_i \int_{-\infty}^{+\infty} dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n f_X(x_1 \ldots x_i \ldots x_n) = \\ &= \int_{-\infty}^{+\infty} dx_i x_i f_{X_i}(x_i) \ .\end{aligned}$$

E' pertanto lecito dire che la componente i-esima della media di $\underline{X}$ è la media della componente i-esima di $\underline{X}$.

Nel caso di una v.c. discreta, ad esempio per una variabile doppia (X, Y) con valori argomentali

$$\begin{aligned}&\arg X = x_1, x_2 \ldots x_r \\ &\arg Y = y_1, y_2 \ldots y_s\end{aligned}$$

si pone

$$\begin{aligned}&(p_{ik} = P(X = x_i, Y = y_k)) \\ &E\{X\} = \sum_{i=1}^{r} \sum_{k=1}^{s} x_i p_{ik} \ , \ E\{Y\} = \sum_{i=1}^{r} \sum_{k=1}^{s} y_k p_{ik} \ ;\end{aligned}$$

ovvero, introducendo le due distribuzioni (discrete) marginali

$$\begin{aligned}p_i &= \sum_{k=1}^{s} p_{ik} = P(X = x_i) \\ q_k &= \sum_{i=1}^{r} p_{ik} = P(Y = y_k) \ ,\end{aligned}$$

$$\begin{cases} E\{X\} = \sum_{i=1}^{r} x_i p_i \\ E\{Y\} = \sum_{k=1}^{s} y_k q_k \ . \end{cases} \tag{14.2}$$

Nel caso poi si abbia una variabile statistica (doppia) non ordinata, formata dall'estrazione di N coppie (x_i, y_i) si porrà

$$\begin{cases} E\{X\} = \frac{1}{N}\sum_{i=1}^{N} x_i \\ E\{Y\} = \frac{1}{N}\sum_{i=1}^{N} y_i \ ; \end{cases} \tag{14.3}$$

ovvia è l'estensione al caso discreto n-dimensionale.

Tornando alle v.c. continue osserviamo che anche nel caso n-dimensionale vale l'importante teorema della media che enunciamo qui senza dimostrazione.

Teorema della media. Sia $\underline{y} = \underline{g}(\underline{x})$ una trasformazione da R_n a R_m; sia $\underline{X}$ una v.c. in R_n ed $\underline{Y}$ la v.c. corrispondente in R_m; posto per definizione

$$E_X\{\underline{g}(\underline{X})\} = \int_{R_n} \underline{g}(\underline{x}) f_X(\underline{x}) dV_n \tag{14.4}$$

si ha, posto che esista la media di Y,

$$E_Y\{\underline{Y}\} = E_X\{\underline{g}(\underline{x})\} \ ^{14}$$

Corollario 14.1: *L'operazione di media è una operazione lineare.* Infatti se

$$\underline{Y} = A\underline{X} + \underline{b}$$

ovvero

$$Y_i = \sum_k a_{ik} X_k + b_i \tag{14.5}$$

allora

[14] Il teorema della media vale per n ed m qualsiasi.

$$\begin{aligned}\mu_{Y_i} &= E_Y\{Y_i\} = E_X\{\sum_k a_{ik}X_k + b_i\} = \\ &= \sum_k a_{ik}E_X\{X_k\} + b_i = \\ &= \sum_k a_{ik}\mu_{X_k} + b_i\end{aligned}$$

ovvero

$$\mu_Y = A\mu_X + \underline{b} \; . \tag{14.6}$$

Corollario 14.2: Se la variabile $\underline{X}$ è ben concentrata in una zona di R_n attorno a μ_X e nella stessa zona la funzione $\underline{y} = \underline{g}(\underline{x})$ è lentamente variabile, allora vale la formula approssimata

$$\mu_Y \cong \underline{g}(\mu_X) \; . \tag{14.7}$$

Per provare questa relazione si passa in primo luogo a linearizzare la $\underline{g}(\underline{x})$:

$$Y_i = g_i(\underline{X}) \cong g_i(\mu_X) + \sum_k \frac{\partial g_i}{\partial x_k}(X_k - \mu_{X_k})$$

ovvero

$$\underline{Y} \cong \underline{g}(\mu_X) + \left[\frac{\partial \underline{g}}{\partial x}\right](\underline{X} - \mu_X) \; . \tag{14.8}$$

Usando poi il fatto che

$$E\{(\underline{X} - \mu_X)\} = 0$$

e ragionando come nel caso monodimensionale si ottiene facilmente la (14.7).

Oltre alla media, si possono definire i momenti di una v.c. n-dimensionale: in generale si ha

$$\mu_{i_1 i_2 \ldots i_k}^{n_1 n_2 \ldots n_k} = E\{x_{i_1}^{n_1} x_{i_2}^{n_2} \ldots x_{i_k}^{n_k}\} \; ; \tag{14.9}$$

analogamente, si definiscono i momenti centrali di X come i corrispondenti momenti della variabile scarto

$$\underline{v} = \underline{X} - \mu_X \; , \tag{14.10}$$
$$\overline{\mu}_{i_1 \ldots i_k}^{n_1 \ldots n_k} = \mu_{i_1 \ldots i_k}^{n_1 \ldots n_k}(v) \; . \tag{14.11}$$

In pratica, di tutti i momenti quelli di gran lunga più usati sono quelli del secondo ordine, che indicheremo

$$c_{ik} = E\{(X_i - \mu_{X_i})(X_k - \mu_{X_k})\} \; . \tag{14.12}$$

Notiamo in primo luogo che per $i = k$ si ottiene

$$c_{ii} = \sigma_i^2 = E\{(X_i - \mu_{X_i})^2\} \tag{14.13}$$

che non è altro che la varianza della componente n-esima, X_i.

Per $i \neq k$ i coefficienti c_{ik}, talvolta anche indicati come σ_{ik}, vengono detti *coefficienti di covarianza* delle componenti X_i ed X_k; si ha ovviamente $c_{ik} = c_{ki}$.

Osserviamo che la (14.12) può essere scritta in forma matriciale

$$\begin{aligned} C_{XX} &= [c_{ik}] = E\{[(X_i - \mu_{X_i})(X_k - \mu_{X_k})]\} = \\ &= E\{(\underline{X} - \mu_X)(\underline{X} - \mu_X)^+\} \; ; \end{aligned} \tag{14.14}$$

la matrice C_{XX} viene appunto detta *matrice di covarianza* della v.c. $\underline{X}$.

C_{XX} è sempre una matrice simmetrica.

Si può anche notare che, essendo

$$(\underline{X} - \mu_X)(\underline{X} - \mu_X)^+ = \underline{X}\underline{X}^+ - \mu_X \underline{X}^+ - \underline{X}{\mu_X}^+ + \mu_X {\mu_X}^+ \; ,$$

e per la linearità della media si ha

$$
\begin{aligned}
C_{XX} &= E\{\underline{X}\underline{X}^+\} - \mu_X {\mu_X}^+ - \mu_X {\mu_X}^+ + \mu_X {\mu_X}^+ = \qquad (14.15)\\
&= E\{\underline{X}\underline{X}^+\} - \mu_X {\mu_X}^+
\end{aligned}
$$

é questa una relazione che generalizza la (8.6).

Osservazione 14.1: supponiamo che le componenti di $\underline{X}$ siano tra loro indipendenti; in tal caso si ha

$$
f_X(\underline{x}) = f_{X_1}(x_1) \dots f_{X_n}(x_n) \ .
$$

Pertanto, osservando che ogni marginale è normalizzata singolarmente,

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

si trova, per $i \neq k$,

$$
\begin{aligned}
E\{X_i X_k\} &= \int_{-\infty}^{+\infty} dx_1 \dots dx_n x_i x_k f_X(x) = \qquad (14.16)\\
&= \int_{-\infty}^{+\infty} dx_i x_i f_{X_i}(x_i) \int_{-\infty}^{+\infty} dx_k x_k f_{X_k}(x_k) = \mu_{X_i} \mu_{X_k} \ .
\end{aligned}
$$

La (14.16) in sostanza dice che gli elementi fuori dalla diagonale di C_{XX} si annullano (vedi (14.15)), così che la matrice di covarianza di n variabili indipendenti assume la caratteristica forma diagonale

$$
C_{XX} = \begin{vmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ & & \ddots \\ 0 & & \sigma_n^2 \end{vmatrix} \qquad (14.17)
$$

Si noti che il viceversa non è vero, cioè la forma diagonale di C_{XX} non significa necessariamente che le componenti di X siano tra loro indipendenti (si veda l'Esempio 14.3 a fine paragrafo).

Osservazione 14.2: come già fatto per la varianza nel caso monodimensionale, vogliamo trovare ora la covarianza di una v.c. $\underline{Y} \in R_m$, funzione di una v.c. $\underline{X} \in R_n$.

Iniziamo con il caso di una funzione lineare

$$\underline{Y} = A\underline{X} + \underline{b} \ :$$

usando la (14.6) si ha

$$\underline{Y} - \mu_Y = A(\underline{X} - \mu_X) \ .$$

Per la (13.14) e ricordando il teorema della media, si ha

$$C_{YY} = E\{(\underline{Y} - \mu_y)(\underline{Y} - \mu_Y)^+\} = E\{A(\underline{X} - \mu_X)(\underline{X} - \mu_X)^+ A^+\} \ .$$

Sfruttando la linearità della media, le matrici A ed A^+ possono essere portate fuori dall'operatore $E\{\bullet\}$, visto che i loro coefficienti sono costanti,

$$C_{YY} = A \ E\{(\underline{X} - \mu_X)(\underline{X} - \mu_X)^+\}A^+ = A \ C_{XX} A^+ \ . \qquad (14.18)$$

La (14.18) è la forma esatta della *legge di propagazione della covarianza* per una trasformazione lineare.

Notiamo che nel caso in cui $m = 1$, cioè

$$\underline{Y} = a_1 X_1 + \ldots + a_n X_n + b \ ,$$

la matrice A è costituita dal semplice vettore

$$A = [a_1 \ldots a_n] = \underline{a}^+ \quad (\underline{a} = \begin{vmatrix} a_1 \\ \vdots \\ a_n \end{vmatrix})$$

mentre la matrice C_{YY} risulta costituita da un solo elemento, σ_Y^2, così che si ha

$$\sigma_Y^2 = \underline{a}^+ C_{XX} \underline{a} = \sum_{i,k=1}^{n} a_i a_k c_{ik} \ . \tag{14.19}$$

La (14.19) è nota come *legge di propagazione degli errori*, nel caso lineare. Se poi le componenti X_i sono variabili tra loro indipendenti, si ha $c_{ik} = 0$ per $i \neq k$, e la (14.19) diventa semplicemente

$$\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 \quad (c_{ii} = \sigma_i^2) \ . \tag{14.20}$$

Questa formula risolve in particolare il problema di conoscere il grado di dispersione di una grandezza Y che sia funzione lineare di altre grandezze misurate indipendentemente da $\{X_i\}$. Poiché le X_i sono quantità misurate, esse saranno descritte da variabili casuali, (indipendenti per ipotesi) di cui si conoscono le varianze, così che la (14.20) ci permette di calcolare la varianza e lo s.q.m. di Y; ovvero l'errore con cui Y è nota attraverso le misure delle $\{X_i\}$ ed i loro errori.

Questo è il motivo per cui le (14.19), (14.20) sono note come legge di propagazione degli errori.

Queste formule possono poi essere generalizzate, in forma approssimata, al caso in cui $\underline{Y}$ non sia una funzione lineare di $\underline{X}$;

$$\underline{Y} = \underline{g}(\underline{X}) \ . \tag{14.21}$$

Se ci mettiamo nell'ipotesi del Corollario 14.2 al teorema della media, la (14.21) può essere linearizzata

$$\underline{Y} \cong \underline{g}(\mu_X) + \left[\frac{\partial \underline{g}}{\partial x}\right] (\underline{X} - \mu_X) \tag{14.22}$$

e, per un ragionamento visto più volte, alla (14.22) si possono applicare le formule del caso lineare. Si ottiene così

$$C_{YY} \cong \left[\frac{\partial \underline{g}}{\partial x}\right] C_{XX} \left[\frac{\partial \underline{g}}{\partial x}\right]^+ \tag{14.23}$$

e nel caso che Y abbia una sola componente

$$\sigma_Y^2 = \sum_{ik} \left(\frac{\partial g}{\partial x_i}\right) \left(\frac{\partial g}{\partial x_k}\right) c_{ik} \tag{14.24}$$

Se $n = 2$, la (14.24) può essere esplicitata nella celebre formula

$$\sigma_Y^2 = \left(\frac{\partial g}{\partial x_1}\right)^2_{\mu_X} \sigma_1^2 + \left(\frac{\partial g}{\partial x_2}\right)^2_{\mu_X} \sigma_2^2 + 2 \left(\frac{\partial g}{\partial x_1}\right)_{\mu_X} \left(\frac{\partial g}{\partial x_2}\right)_{\mu_X} \sigma_{12} \ . \tag{14.25}$$

E' bene osservare che tutte le derivate dalla formula (14.22) alla (14.25) vanno considerate come calcolate in μ_X o quanto meno in un punto "vicino" a μ_X.

Osservazione 14.3: terminiamo la teoria di questo paragrafo definendo una proprietà assai importante delle matrici di covarianza.

La matrice C_{XX} è sempre definita positiva, cioè

$$\underline{a}^+ C_{XX} \underline{a} \geq 0 \quad \forall \underline{a} \in R_n \ ; \tag{14.26}$$

se inoltre X non ha una distribuzione singolare, C_{XX} è definita positiva in senso stretto, cioè

$$\underline{a}^+ C_{XX} \underline{a} > 0 \quad \forall \underline{a} \neq 0 \in R_n \ ; \tag{14.27}$$

Infatti, si fissi un qualunque vettore costante $\underline{a}$ e si consideri la variabile

$$Y = \underline{a}^+ \underline{X} \ ;$$

per la (14.19)

$$\underline{a}^+ C_{XX} \underline{a} = \sigma_Y^2 = E\{[\underline{a}^+ (\underline{X} - \mu_X)]^2\} \geq 0 \ :$$

la (14.26) è così provata.

Inoltre, notiamo che se per una qualche $\underline{a}$ si avesse

$$\underline{a}^+ C_{XX} \underline{a} = 0$$

allora sarebbe anche

$$E\{[\underline{a}^+(\underline{X}-\mu_X)]^2\}=0\ . \tag{14.28}$$

Ma essendo $\{\underline{a}^+(\underline{X}-\mu_X)\}^2$ una variabile casuale (funzione della $\underline{X}$) sempre positiva, la (14.28) può essere verificata se e solo se

$$\underline{a}^+(\underline{X}-\mu_X)=0 \tag{14.29}$$

con probabilità $P=1$: ma la (14.29) significa che $\underline{X}$ è una variabile distribuita in R_n su un piano a $n-1$ dimensioni passante per μ_X e perpendicolare ad $\underline{a}$, cioè il piano di equazione $a_1x_1+\ldots+a_nx_n-a_1\mu_{X_1}-\ldots-a_n\mu_{X_n}=0$. Poiché abbiamo escluso per ipotesi che $\underline{X}$ sia singolare, la (14.29) non può essere verificata e quindi

$$\underline{a}^+C_{XX}\underline{a}>0 \quad , \quad \forall\underline{a}\neq 0\ .$$

Notiamo che in particolare la (14.27) comporta anche che

$$\det C_{XX}\neq 0 \tag{14.30}$$

e quindi la matrice C_{XX} è regolare, cioè invertibile.

Notiamo infine che essendo C_{XX} una matrice simmetrica e definita positiva, è sempre possibile trovare un'altra matrice K simmetrica e definita positiva, tale che[15]

$$K^2=C_{XX}\ ; \tag{14.31}$$

K è detta la radice positiva di C_{XX}.

[15]Più precisamente si ha che C_{XX} può essere espressa come

$$C_{XX}=U\Lambda U^+$$

dove Λ è la matrice diagonale degli autovalori di C_{XX}, ed U è una matrice ortogonale ($U^+U=I$): pertanto è facile vedere che

$$K=U\Lambda^{1/2}U^+$$

essendo $\Lambda^{1/2}$ la matrice diagonale costruita con le radici degli autovalori di C_{XX}.

Osservazione 14.4: oltre ai momenti di ordine 1 e 2, usati per definire media e covarianza, si possono naturalmente definire momenti e momenti centrali di ordine superiore.

Oltre a ciò si può anche definire una funzione generatrice dei momenti, per quelle distribuzioni per cui esiste, ovvero

$$\begin{aligned} &G_{\underline{X}}(\underline{t}) = E\{e^{\underline{t}^+ \underline{X}}\} \\ &(\underline{t}^+ \underline{X} = t_1 X_1 + t_2 X_2 + \ldots t_n X_n) \ . \end{aligned} \tag{14.32}$$

Qualora la (14.32) esista in una sfera $|\underline{t}| \leq r$ attorno all'origine dello spazio dei parametri, la $G_{\underline{X}}(\underline{t})$ può essere usata per costruire i momenti secondo la regola

$$\frac{\partial^{\alpha_1+\alpha_2\ldots+\alpha_n}}{\partial t_1^{\alpha_1} \ldots \partial t_n^{\alpha_n}} G_{\underline{X}}(\underline{t})|_{\underline{t}=0} = E\{X_1^{\alpha_1} X_2^{\alpha_2} \ldots X_n^{\alpha_n}\} \ . \tag{14.33}$$

Esempio 14.1: si verifichi il teorema della media nel caso (vedi la formula (13.18))

$$\begin{aligned} &(X,Y) \ : && f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \\ &\xi = X^2 + Y^2 \ : && f_\xi(\xi) = \frac{1}{2} e^{-\frac{\xi}{2}} \qquad\qquad (\xi \geq 0) \end{aligned} \ .$$

Si ha

$$\mu_\xi = E\{\xi\} = \int_{-\infty}^{+\infty} d\xi \ \xi \frac{1}{2} e^{-\frac{\xi}{2}} = 2 \int_{-\infty}^{+\infty} d\eta \ \eta e^{-\eta} = 2 \ .$$

Inoltre

$$E_{(X,Y)}\{X^2 + Y^2\} = E\{X^2\} + E\{Y^2\} \ ;$$

poiché (X,Y) sono indipendenti e ciascuna distribuita secondo una normale standardizzata

$$\mu_X = \mu_Y = 0$$
$$E\{X^2\} = E\{Y^2\} = 1 \ .$$

Pertanto risulta

$$\mu_\xi = 2 = E_{(X,Y)}\{X^2 + Y^2\}$$

e il teorema della media è verificato.

Esempio 14.2: siano X_1 ed X_2 due variabili normali indipendenti

$$X_1 = N[\mu_1, \sigma_1^2]$$
$$X_2 = N[\mu_2, \sigma_2^2]$$
$$f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \ .$$

Vogliamo mostrare che la somma

$$Y = X_1 + X_2$$

é ancora una variabile normale.

Osserviamo in primo luogo che per le (14.6) e (14.20)

$$\mu_Y = \mu_1 + \mu_2$$
$$\sigma_Y^2 = \sigma_1^2 + \sigma_2^2 \ .$$

Per vedere che Y è normale, osserviamo che, usando la definizione di funzione generatrice dei momenti ed il teorema della media, si ha

$$\begin{aligned} G_Y(t) &= E\{e^{tY}\} = E\{e^{t(X_1+X_2)}\} = E\{e^{tX_1} e^{tX_2}\} = \\ &= E\{e^{tX_1}\} E\{e^{tX_2}\} = G_{X_1}(t) G_{X_2}(t) \ . \end{aligned}$$

La possibilità di porre la media del prodotto $e^{tX_1} e^{tX_2}$ uguale al prodotto delle medie ci è data dalla indipendenza di X_1 e X_2, per cui

$$
\begin{aligned}
E\{e^{tX_1}e^{tX_2}\} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx_1 dx_2 e^{tx_1}e^{tx_2} f_{X_1}(x_1) f_{X_2}(x_2) = \\
&= \int_{-\infty}^{+\infty} dx_1 e^{tx_1} f_{X_1}(x_1) \int_{-\infty}^{+\infty} dx_2 e^{tx_2} f_{X_2}(x_2) = \\
&= E\{e^{tX_1}\}E\{e^{tX_2}\} \ .
\end{aligned}
$$

Ricordando la (10.6) si trova

$$
\begin{aligned}
G_Y(t) &= e^{\mu_1 t} e^{\frac{1}{2}\sigma_1^2 t^2} e^{\mu_2 t} e^{\frac{1}{2}\sigma_2^2 t^2} = \\
&= e^{(\mu_1+\mu_2)t} e^{\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2} \ ;
\end{aligned}
$$

poiché la funzione generatrice dei momenti è in corrispondenza biunivoca con la funzione di densità di probabilità di una v.c., questa relazione conferma che Y è una variabile normale con media $\mu_1 + \mu_2$ e varianza $\sigma_1^2 + \sigma_2^2$.

Osservazione 14.5: questo esempio è facilmente generalizzabile al caso della somma di n variabili normali indipendenti

$$
\begin{cases}
X_i = N[\mu_i, \sigma_i^2] \qquad i = 1, 2 \ldots n \\
f_X(x_1 \ldots x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n)
\end{cases}
\tag{14.34}
$$

$$
\begin{cases}
Y = \displaystyle\sum_{i=1}^{n} X_i \\
\\
Y = N[\Sigma_i \mu_i, \Sigma_i \sigma_i^2] \ .
\end{cases}
$$

Il teorema di somma delle variabili normali è in realtà valido in generale, anche se le X_i non sono tra loro indipendenti; in tal caso però occorre usare la (14.19) invece della (14.20) per il calcolo di σ_Y^2.

Esempio 14.3: si calcoli media e covarianza della distribuzione bidimensionale (vedi Esempio §12)

$$f(x,y) = \begin{cases} \dfrac{3}{2\pi R^3}\sqrt{R^2 - x^2 - y^2} & (x^2 + y^2 \leq R^2) \\ 0 & (x^2 + y^2 > R^2) \ . \end{cases}$$

Per simmetria è evidente che risulta

$$\mu_X = \mu_Y = 0 \ .$$

In effetti

$$\mu_X = \int_{-\infty}^{+\infty} dx \ x f_X(x) = \int_{-R}^{R} x \frac{3}{4R^3}(R^2 - x^2)dx = 0$$

e analogamente μ_Y.
Quanto alla covarianza osserviamo che

$$C = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} = \begin{vmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{vmatrix} \ ;$$

$$\begin{aligned} \sigma_X^2 &= E\{X^2\} = \frac{3}{4R^3} \cdot 2 \int_0^R x^2(R^2 - x^2)dx = \frac{R^2}{5} \\ \sigma_Y^2 &= E\{Y^2\} = \frac{R^2}{5} \\ \sigma_{XY} &= E\{XY\} = \frac{3}{2\pi R^3} \int_{-R}^{R} dx \ x \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy \ y\sqrt{R^2 - x^2 - y^2} = 0 \ . \end{aligned}$$

Pertanto

$$C = \frac{R^2}{5} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \ :$$

notiamo che la covarianza del vettore $\begin{vmatrix} X \\ Y \end{vmatrix}$ é diagonale senza che perciò X e Y siano indipendenti tra loro.

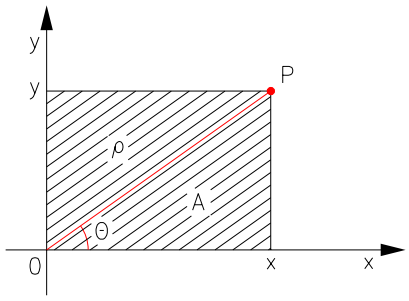


Figura 14.1

Esempio 14.4: di un punto P si sono misurate la distanza dall'origine ρ, e l'anomalia θ (fig. 14.1).

Sapendo che le misure (ρ, θ) sono indipendenti e che sono rappresentate da variabili con media

$$\mu(\rho) = 1 \text{ km}$$
$$\mu(\theta) = \pi/6$$

e con s.q.m.

$$\sigma(\rho) = 1 \text{ mm}$$
$$\sigma(\theta) = 2.10^{-6} \text{ (rad.)}$$

vogliamo conoscere media e covarianza delle coordinate (X, Y) di P e media e varianza dell'area A del rettangolo che ha OP come diagonale. Indichiamo con

$$\eta = \begin{vmatrix} X \\ Y \end{vmatrix} \ , \ \xi = \begin{vmatrix} \rho \\ \theta \end{vmatrix} \ .$$

Per ipotesi la matrice di covarianza di ξ è data da

$$C_{\xi\xi} = \begin{vmatrix} \sigma_\rho^2 & 0 \\ 0 & \sigma_\theta^2 \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 4.10^{-12} \end{vmatrix} .$$

Notiamo che ponendo $\sigma_\rho^2 = 1$ conveniamo di misurare tutte le lunghezze in mm, a meno di specifiche indicazioni.

La relazione tra η e ξ è data da

$$\eta = \begin{vmatrix} x \\ y \end{vmatrix} = g(\xi) = \begin{vmatrix} \rho\cos\theta \\ \rho\sin\theta \end{vmatrix} .$$

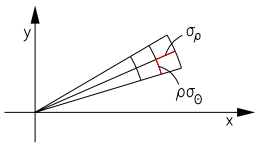
Poiché ξ è una variabile assai concentrata possiamo porre[16]

$$\mu_X = \mu_\rho \cos\mu_\theta = 866025 \ \text{mm}$$
$$\mu_X = \mu_\rho \sin\mu_\theta = 500000 \ \text{mm} .$$

Inoltre lo Jacobiano di η rispetto a ξ è dato da

$$\left|\frac{\partial g}{\partial \xi}\right| = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix}$$

[16]Poiché $\sigma = 1$ mm e $\rho\sigma_\theta = 2$ mm, è chiaro che in un'area di pochi mm^2 si ha una alta probabilità di individuare P.



Allora, per rendere più precisa l'affermazione possiamo calcolare per la X il resto del secondo ordine trascurato con la linearizzazione: indicati con $\overline{\rho}, \overline{\theta}$ i valori medi di ρ e θ si ha:

$$x = \overline{\rho}\cos\overline{\theta} + \cos\overline{\theta}(\rho - \overline{\rho}) - \overline{\rho} \ \ (\sin\overline{\theta}(\theta - \overline{\theta}) +$$
$$+\frac{1}{2}[-2\sin\overline{\theta}(\theta - \overline{\theta})(\rho - \overline{\rho}) - \overline{\rho}\cos\overline{\theta}(\theta - \overline{\theta})^2] + \ldots$$

Come si vede, il termine quadratico è dell'ordine di 10^{-6} mm e quindi anche la sua media, nell'area dove è concentrata la probabilità, è assolutamente trascurabile.

così che

$$\left|\frac{\partial g}{\partial \xi}\right|_{\mu_\xi} = \begin{vmatrix} 0,866 & -10^6 \cdot 0,500 \\ 0,500 & 10^6 \cdot 0,866 \end{vmatrix}$$

ed in particolare

$$\det \left|\frac{\partial g}{\partial \xi}\right|_{\mu_\xi} = \overline{\rho}(\cos^2 \overline{\theta} + \sin^2 \overline{\theta}) = \overline{\rho} \neq 0 \ .$$

Applicando la (14.23) si trova allora

$$\begin{aligned} C_{\eta\eta} &= \begin{vmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_y^2 \end{vmatrix} = \\ &= \begin{vmatrix} 0,866 & -10^6 \cdot 0,500 \\ 0,500 & 10^6 \cdot 0,866 \end{vmatrix} \begin{vmatrix} 1 & 0 \\ 0 & 4.10^{-12} \end{vmatrix} \cdot \\ &\cdot \begin{vmatrix} 0,866 & 0,500 \\ -10^6 \cdot 0,500 & 10^6 \cdot 0,866 \end{vmatrix} = \begin{vmatrix} 1,750 & -1,299 \\ -1,299 & 3,250 \end{vmatrix} \ . \end{aligned}$$

E' interessante notare che mentre la covarianza di ρ e θ é nulla, così non è per la covarianza di X ed Y che è diversa da zero: cioè ρ e θ sono indipendenti stocasticamente per ipotesi, mentre X e Y sono tra loro stocasticamente dipendenti.

Passiamo ora a considerare l'area del rettangolo di fig. 14.1: si ha

$$A = \rho \sin\theta \rho \cos\theta = \frac{1}{2}\rho^2 \sin 2\theta \ .$$

La media di A sarà allora data da

$$\mu(A) = 10^{12} \cdot 0,433013 \text{ mm}^2 = 10^6 \cdot 0,433013 \text{ m}^2 \ .$$

Per la varianza di A useremo la (14.25): notando che $\sigma_{\rho\theta} = 0$, risulta

$$\begin{aligned}\sigma^2(A) &= (\overline{\rho}\sin 2\overline{\theta})^2\sigma_\rho^2 + (\overline{\rho}^2\cos 2\overline{\theta})^2\sigma_\theta^2 = \\ &= 10^{12}\cdot\frac{3}{4}\cdot 1 + 10^{24}\cdot\frac{1}{4}\cdot 4\cdot 10^{-12} = \\ &= 1,75\cdot 10^{12}(\mathrm{mm}^2)^2\end{aligned}$$

ovvero

$$\sigma(A) = 1,323\cdot 10^6\ \mathrm{mm}^2 = 1,323\ \mathrm{m}^2\ .$$

15 La dipendenza funzionale tra le componenti di una variabile casuale: l'indice di Pearson

In questo paragrafo per semplicità considereremo una v.c. doppia di componenti (X, Y).

Per una tale variabile sono già state definite due situazioni estreme:

a) Y è stocasticamente dipendente da X; in questo caso

$$f(x,y) = f_X(x)f_Y(y)\ ; \tag{15.1}$$

b) la variabile doppia (X, Y) è singolare e quindi esiste una curva L, che supponiamo abbia equazione

$$y = g(x)\ ,$$

tale che risulti

$$P((X,Y)\in L) = 1\ ;$$

in questo caso diremo che Y è funzionalmente dipendente da X:

Ogni v.c. doppia che non rientri né nel caso a) né nel caso b) costituisce una situazione intermedia; di tale situazione è però possibile dire se essa

si avvicini maggiormente al caso a) o al caso b). Questa misura può essere effettuata in diversi modi; in questo paragrafo seguiremo la via della definizione di curva di regressione, curva di variabilità ed indice di Pearson.

In sostanza si può osservare che nel caso b), fissata X la distribuzione di probabilità della Y condizionata al valore di X è data semplicemente da

$$\begin{cases} P\{Y = g(x)|X = x\} = 1 \\ P\{Y \neq g(x)|X = x\} = 0 \ ; \end{cases}$$

ne deriva quindi che, sempre fissata $X = x$, la media e la varianza condizionata di Y sono

$$E\{Y|X = x\} = g(x) \tag{15.2}$$
$$\sigma^2(Y|X = x) = 0 \ . \tag{15.3}$$

E' chiaro che quanto più ci si avvicina alle condizioni (15.2), (15.3) tanto più la v.c. doppia in considerazione è vicina ad una situazione di tipo b)

Per questo motivo si definiscono:

la curva di regressione di Y su X:

$$\begin{aligned} \overline{y}(x) = E\{Y|X = x\} &= \int_{-\infty}^{+\infty} y \ f_{Y|X}(y|x) \ dy = \\ &= \int_{-\infty}^{+\infty} y \frac{f(x,y)}{f_X(x)} \ dy \ ; \end{aligned} \tag{15.4}$$

la curva di variabilità di Y attorno ad $\overline{y}(x)$:

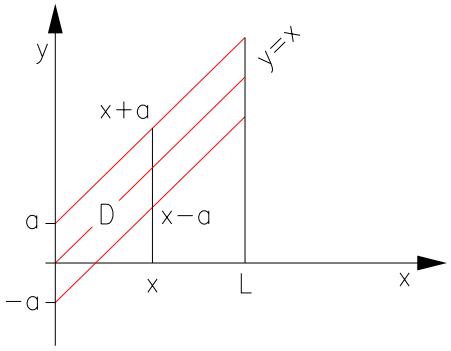
$$\begin{aligned} \sigma^2(Y|X = x) &= E\{(Y - \overline{y}(x))^2|X = x\} = \\ &= \int_{-\infty}^{+\infty} (y - \overline{y}(x))^2 f_{Y|X}(y|x) dy \ . \end{aligned} \tag{15.5}$$

E' chiaro che più piccola sarà la curva di variabilità, più la variabile doppia tenderà a concentrarsi sulla curva $y = \overline{y}(x)$. Per poter disporre di un indice globale si introduce la *varianza residua*

$$\begin{aligned}
\sigma_R^2 &= \int_{-\infty}^{+\infty} \sigma^2(Y|X=x) f_X(x)\, dx = \\
&= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy (y - \overline{y}(x))^2 f_{Y|X}(y|x) f_X(x) = \qquad (15.6) \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx\, dy (y - \overline{y}(x))^2 \;\; f(x,y) = E\{(Y - \overline{y}(X))^2\} \quad :
\end{aligned}$$

questo indice misura la *dispersione globale* della variabile doppia rispetto alla curva di regressione $\overline{y}(x)$.

Esempio 15.1: si consideri la distribuzione uniforme nel dominio in fig. 15.1.



$$f(x,y) = 1/2aL \text{ in } D$$

Figura 15.1

E'chiaro che per $a \to 0$ la v.c. tende ad una variabile singolare distribuita uniformemente sul segmento $\{y = x; 0 \le x \le L\}$.

La curva di regressione, per motivi di simmetria, è data da

$$\overline{y}(x) = x$$

mentre la varianza residua è

$$\sigma_R^2 = \int_0^L dx \int_{x-a}^{x+a} dy (y-x)^2 \frac{1}{2aL} = \frac{1}{L} \int_0^L dx \frac{1}{3} a^2 = \frac{1}{3} a^2 \; .$$

Come si vede $\sigma_R \to 0$ implica $a^2 \to 0$ e quindi la tendenza di Y ad essere funzionalmente dipendente da X.

Osservazione 15.1: per come è stata definita la curva di regressione, $\overline{y}(x)$ indica la dipendenza di una variabile ordinaria $\overline{y}$ da un'altra variabile ordinaria x. Tuttavia, una volta definita $\overline{y}(x)$ come funzione ordinaria, è possibile considerare la v.c. $\overline{y}(X)$ indicata col simbolo $\overline{y}(X) = E\{Y|X\}$.

Tale variabile, che è per definizione funzione della sola X, si chiama la variabile Y condizionata alla X.

Osservazione 15.2: notiamo che se $\varphi(X,Y)$ è una qualsiasi funzione della v.c. (X,Y), vale la relazione

$$E\{\varphi(X,Y)\} = E_X\left\{E_Y\{\varphi(X,Y)|X\}\right\} \ ; \tag{15.7}$$

infatti, ricordando che $f_{Y|X}(y|x)f_X(x) = f(x,y)$, si ha

$$\begin{aligned} E_X\{E_Y\{\varphi(X,Y)|X\}\} &= \int_{-\infty}^{+\infty} dx\ f_X(x) \int_{-\infty}^{+\infty} dy\ \varphi(x,y) f_{Y|X}(y|x) = \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx\ dy\ \varphi(x,y)\ f(x,y) = \\ &= E\{\varphi(X,Y)\}\ . \end{aligned}$$

In particolare dalla (15.7) si deduce che se $\varphi(X)$ è una qualunque funzione della sola X allora

$$E\{\phi(X)(Y - \overline{y}(X))\} = 0\ . \tag{15.8}$$

Infatti

$$\begin{aligned} E\{\varphi(X)(Y-\overline{y}(X))\} &= E_X\{E_Y\{\varphi(X)(Y-\overline{y}(X))|X\}\} = \\ &= E_X\{\varphi(X)E_Y\{Y-\overline{y}(X)|X\}\} = \\ &= E_X\{\varphi(X)\{E_Y\{Y|X\} - \overline{y}(X)\}\} = 0\ . \end{aligned}$$

Usando la (15.8) risulta assai semplice provare il seguente teorema.

Teorema 15.1: la variabile di regressione $\overline{Y} = \overline{y}(X)$ è tra tutte le variabili del tipo [17] $Y = g(X)$ quella che meglio "spiega" il legame tra Y ed X nel senso che

$$E\{(Y - g(X))^2\} \geq E\{(Y - \overline{y}(X))^2\} = \sigma_R^2 \ . \qquad (15.9)$$

Infatti, per la (15.8) si ha

$$\begin{aligned} & E\{(Y - g(X))^2\} = E\{(Y - \overline{y}(X) + \overline{y}(X) - g(X))^2\} \qquad (15.10) \\ & = E\{(Y - \overline{y}(X))^2\} + 2E\{(Y - \overline{y}(X))(\overline{y}(X) - g(X))\} + \\ & + E\{(\overline{y}(X) - g(X))^2\} = \\ & = \sigma_R^2 + E\{(\overline{y}(X) - g(X))^2\} \ , \end{aligned}$$

da cui ovviamente discende la (15.9) .

Notiamo poi che se si sceglie $g(x) = \mu_Y$, la (15.10) assume un particolare significato di decomposizione degli scarti della Y.

$$\sigma_Y^2 = E\{(Y - \mu_Y)^2\} = \sigma_R^2 + E\{(\overline{y}(X) - \mu_Y)^2\} \ : \qquad (15.11)$$

ovvero la variabilità della Y attorno a μ_Y , misurata da σ_Y^2, deriva da una variabilità "spiegata" dal legame tra Y e X, misurata da $\sigma_S^2 = E\{(\overline{y}(X) - \mu_Y)^2\}$, che prende il nome di varianza spiegata, cui va aggiunta la variabilità di Y attorno alla regressione, misurata dalla varianza residua σ_R^2.

La (15.11) perciò può semplicemente essere scritta

$$\sigma_Y^2 = \sigma_R^2 + \sigma_S^2 \ ; \qquad (15.12)$$

questa relazione ha suggerito al Pearson di assumere come misura della dipendenza della Y da X l'indice

$$\eta_Y^2 = \frac{\sigma_S^2}{\sigma_Y^2} \ . \qquad (15.13)$$

[17] Con la sola restrizione che la v.c. $Y - g(X)$, funzione di (X, Y), ammetta media e varianza.

Dalla (15.12) derivano alcune evidenti proprietà dell'indice di Pearson:

a) η^2_Y è sempre compreso tra 0 e 1;

b) $\eta^2_Y = 1$ se e solo se Y è funzionalmente dipendente da X; infatti $\eta^2_Y = 1$ implica $\sigma^2_R = 0$;

c) $\eta^2_Y = 0$ se Y è indipendente da X, ma non è detto il viceversa; infatti $\eta^2_Y = 0$ implica $\sigma^2_S = 0$ cioè $\overline{y}(x) = \mu_Y$, che è certamente verificata se Y é indipendente da X, ma può essere verificata anche in altri casi (vedi l'Esempio 15.2).

Osservazione 15.3: in base ai valori di η^2_Y diremo orientativamente che la dipendenza di Y da X è debole per $\eta^2_Y < 0,4$, consistente per $0,4 < \eta^2_Y < 0,6$, forte se $0,6 < \eta^2_Y < 0,8$, fortissima se $\eta^2_Y > 0,8$.

Osservazione 15.4: quanto detto finora sulla dipendenza di Y da X può anche essere applicato allo studio della dipendenza di X da Y.

Potremo perciò definire una curva di regressione $\overline{x}(y)$ e variabilità - $\sigma^2(X|y)$ di X su Y ed un indice di Pearson η^2_X; queste quantità risultano in generale diverse dalle corrispondenti quantità per la Y.

Esempio 15.2: come già visto nell'Esempio al §12, la distribuzione

$$f(x,y) = \begin{cases} \frac{3}{2\pi R^3}\sqrt{R^2 - x^2 - y^2} & (x^2 + y^2 \leq R^2) \\ 0 & (x^2 + y^2 > R^2) \end{cases}$$

ha come distribuzione della Y condizionata alla X

$$f_{Y|X}(y|x) = \frac{2}{\pi}\frac{\sqrt{R^2 - x^2 - y^2}}{R^2 - x^2} \qquad (|y| \leq \sqrt{R^2 - x^2}) \ .$$

Pertanto si ha

$$\begin{aligned} \overline{y}(x) &= E\{Y|X = x\} = \\ &= \frac{2}{\pi}\frac{1}{R^2 - x^2}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy \ y\sqrt{R^2 - x^2 - y^2} = 0 \end{aligned}$$

$$\begin{aligned}\sigma^2(Y|X=x) &= E\{Y^2|X=x\} = \\ &= \frac{2}{\pi}\frac{1}{R^2-x^2}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy\; y^2\sqrt{R^2-x^2-y^2} = \\ &= \frac{1}{4}(R^2-x^2)\ .\end{aligned}$$

Notiamo che in questo caso

$$\overline{y}(x) = \mu_Y = 0$$

così che l'indice di Pearson risulta pure esso

$$\eta_Y^2 = 0\ ,$$

sebbene, come si è più volte detto, la Y non sia stocasticamente indipendente da X.

Esempio 15.3: vediamo in questo esempio come si applicano le formule di questo paragrafo ad una variabile discreta. La variabile che consideriamo è descritta dalla seguente tabella (nei riquadri vuoti si sottintenda una probabilità nulla).

x_i \ y_k	2	4	6	8	p_i
1	0,1	0,1			0,2
2	0,1	0,1	0,1		0,3
3		0,1	0,1	0,1	0,3
4			0,1	0,1	0,2
q	0,2	0,3	0,3	0,2	1,0

Le distribuzioni condizionate della Y sono date da

$$P(Y=y_k|X=x_i) = \frac{p_{ik}}{p_i}$$

x_i \ y_k	2	4	6	8
1	0,50	0,50		
2	0,33	0,33	0,33	
3		0,33	0,33	0,33
4			0,50	0,50

Pertanto la curva di regressione $\overline{y}(x_i)$ è data da:

$$\overline{y}(x_i) = E\{Y|x_i\} = \sum_k \frac{p_{ik}}{p_i} y_k$$

x_i	$\overline{y}(x_i)$
1	3
2	4
3	6
4	7

mentre la curva di variabilità è data da:

$$\sigma^2(Y|x_i) = \sum_k (y_k - \overline{y}(x_i))^2 \frac{p_{ik}}{p_i}$$

x_i	$\sigma^2(y|x_i)$
1	1
2	2,67
3	2,67
4	1

Ora notiamo che

$$\begin{aligned} \mu_Y &= \sum_{ik} y_k p_{ik} = \sum_i p_i \sum_k y_k \frac{p_{ik}}{p_i} = \\ &= \sum_i p_i \overline{y}(x_i) = 5 \ . \end{aligned}$$

E' facile perciò calcolare

$$\sigma_R^2 = \sum_i \sigma^2(Y|x_i) p_i = 2$$
$$\sigma_S^2 = \sum_i (\overline{y}(x_i) - \mu_Y)^2 p_i = 2,2 \ .$$

Usando le (15.12) e (15.13) si trova quindi

$$\eta_Y^2 = \frac{\sigma_S^2}{\sigma_Y^2} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_R^2} = 0,52 \ ;$$

come si vede (e come era facilmente arguibile), la Y dipende sensibilmente dalla X.

16 L'indice di correlazione lineare

Anche in questo paragrafo proseguiamo l'analisi della dipendenza fra due componenti di una v.c. introducendo un indice particolarmente adatto a rilevare una dipendenza lineare tra di esse. Anche qui svilupperemo il caso bidimensionale essendo ovvia la generalizzazione a n dimensioni.

Iniziamo con l'osservare che se X ed Y sono due variabili indipendenti si ha

$$\sigma_{XY} = E\{XY\} - \mu_X \mu_Y = 0 \ . \tag{16.1}$$

All'estremo opposto supponiamo che la Y sia linearmente (funzionalmente) dipendente da X, cioè

$$Y = aX + b$$

da cui segue

$$\begin{aligned} &Y - \mu_Y = a(X - \mu_X) \\ &\sigma_{XY} = E\{(X - \mu_X)(Y - \mu_Y)\} = a \ E\{(X - \mu_X)^2\} = a\sigma_X^2 \end{aligned} \tag{16.2}$$

Notando che

$$\sigma_Y^2 = a^2 \sigma_X^2 \ , \ \sigma_Y = |a| \sigma_X \ ,$$

la (16.2) può anche essere riscritta

$$\sigma_{XY} = \frac{a}{|a|}\sigma_X\sigma_Y$$

ovvero

$$\frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \pm 1$$

a seconda che a sia positivo o negativo.

Questa semplice analisi ci induce a considerare come indice di dipendenza lineare tra Y e X la quantità

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \ , \tag{16.3}$$

detta appunto indice di correlazione lineare di Y ed X.

Vediamo alcune importanti proprietà di questo indice.

a) Un primo evidente vantaggio è che ρ è invariante in modulo per trasformazioni lineari individuali di X ed Y: infatti

$$\begin{aligned}
&\xi = a_1 x + b_1 \ , \xi - \mu_\xi = a_1(x - \mu_X) \\
&\eta = a_2 y + b_2 \ , \ \eta - \mu_\eta = a_2(y - \mu_Y) \\
&\sigma_{\xi\eta} = a_1 a_2 \sigma_{XY} \\
&\sigma_\xi = |a_1|\sigma_X \ , \ \sigma_\eta = |a_2|\sigma_Y \\
&\rho_{\xi\eta} = \frac{a_1 a_2}{|a_1 a_2|}\rho_{XY} = \pm\rho_{XY} \ .
\end{aligned}$$

Questo comporta in particolare che ρ non cambia se si cambiano le unità di misura in cui sono espresse le quantità X, Y.

b) Per quanto già visto all'inizio del paragrafo, se X e Y sono indipendenti, $\rho = 0$; se sono linearmente dipendenti $\rho = \pm 1$; ci proponiamo ora di provare che in ogni caso

$$-1 \leq \rho \leq 1 \tag{16.4}$$

ed inoltre

$$\rho = \pm 1 \tag{16.5}$$

se e solo se

$$Y = aX + b \ ,$$

ed in particolare $\rho = +1$ se a è positivo, $\rho = -1$ se a é negativo.
Infatti poniamo per comodità

$$v_X = X - \mu_X \ , \ v_Y = Y - \mu_Y \ ,$$

si ha, qualunque sia il valore del parametro α,

$$\begin{aligned} E\{(v_Y - \alpha v_X)^2\} &= E\{v_Y^2 - 2\alpha v_Y v_X + \alpha^2 v_X^2\} = \\ &= \sigma_Y^2 - 2\alpha\sigma_{XY} + \alpha^2\sigma_X^2 \geq 0 \ . \end{aligned} \tag{16.6}$$

Se ora poniamo $\alpha = \dfrac{\sigma_{XY}}{\sigma_X^2}$ nella (16.6), si trova

$$\sigma_Y^2 - 2\frac{\sigma_{XY}^2}{\sigma_X^2} + \frac{\sigma_{XY}^2}{\sigma_X^2} \geq 0$$

cioè

$$\rho_{XY}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \leq 1 \ ,$$

che appunto prova la (16.4).
Inoltre osserviamo che

$$\rho_{XY}^2 = 1 \qquad (\text{cioè } \rho_{XY} = \pm 1)$$

comporta anche (vedi (16.6))

$$E\{(v_Y - \frac{\sigma_{XY}^2}{\sigma_X^2} v_X)^2\} = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} = 0 \ : \qquad (16.7)$$

ma la (16.7) appunto ci dice che

$$v_Y - \frac{\sigma_{XY}}{\sigma_X^2} v_X = Y - \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}(X - \mu_X) = 0 \qquad (16.8)$$

con probabilità $P = 1$, cioè Y è linearmente dipendente da X.

Notiamo che, essendo per ipotesi

$$\sigma_{XY} = \pm \sigma_X \sigma_Y$$

la (16.8) può essere riscritta nella forma

$$\frac{Y - \mu_Y}{\sigma_Y} = \pm \frac{X - \mu_X}{\sigma_X}$$

che è facilmente memorizzabile.

Osservazione 16.1: se il coefficiente di correlazione lineare di X e Y è nullo, si dice che le due variabili sono tra loro incorrelate.

Per quanto detto è ovvio che se X ed Y sono indipendenti, allora sono anche incorrelate, mentre in generale non è vero il viceversa: si veda come controesempio l'Esempio 14.3, in cui X ed Y non sono indipendenti e tuttavia $\sigma_{XY} = 0$ e quindi anche $\rho_{XY} = 0$.

Osservazione 16.2: l'espressione (16.3) è naturalmente valida sia per v.c. continue che discrete. In particolare, nel caso di una variabile statistica doppia non ordinata si preferisce usare la forma equivalente

$$\rho_{XY} = \frac{N \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[N \sum x_i^2 - (\sum x_i)^2][N \sum y_i^2 - (\sum y_i)^2]}} \ , \qquad (16.9)$$

conveniente dal punto di vista del calcolo.

Esempio 16.1: dalla variabile casuale (X, Y)

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2)}$$

si sono estratte le 50 coppie di valori della tabella riportata di seguito.

Poiché X e Y sono tra loro indipendenti, sappiamo che

$$\rho_{XY} = 0 \ :$$

vogliamo verificare, usando la (16.9), che anche per la variabile statistica (X_i, Y_i) si ha all'incirca $\rho = 0$;

$$\begin{array}{lll} \sum x_i = & -12.266 & \\ \sum x_i^2 = & 46.924 & \\ & \sum x_i y_i = & -2.611 \\ & \rho_{XY} = & -0.040 \end{array} \qquad \begin{array}{ll} & \\ \sum y_i = & 1.425 \\ \sum y_i^2 = & 73.252 \\ & \\ & \end{array}$$

che è appunto un valore assai piccolo.

x_i	y_i	x_i	y_i
-0,098	-0,465	0,246	-0,742
0,437	-2,120	-0,803	0,574
-1,812	-2,748	-1,995	0,703
-0,999	0,308	0,313	0,220
-0,852	0,118	-1,372	-1,896
0,697	-1,787	-2,290	-0,616
1.200	0,063	0,053	-2,034
-0,580	0,377	0,659	-0,298
-0,393	-1,412	-1,675	1,1090
-0,404	1,322	0,912	-0,710
0,729	0,204	-0,604	0,966
0,393	-3,090	-0,716	-0,482
0,002	0,261	-1,276	0,653
1,739	0,476	0,340	1,498
-0,043	2,290	-1,655	0,999
-1,616	0,334	0,083	0,824
-0,103	1,041	0,215	-0,426
-0,574	1,131	0,123	2,748
0,671	0,852	0,256	0,298
0,556	1,112	0,138	-0,838
-0.755	0,329	0,755	-0,459
0,454	-0,118	-1,311	-0,703
0,550	-1,254	-0,448	1,015
-0,340	1,655	0,073	-0,762
-1,838	-0,410	-0,678	1,335

17 La distribuzione normale a n dimensioni

Chiamiamo normale (o gaussiana) una qualsiasi v.c. distribuita secondo la legge

$$f(\underline{x}) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} e^{-1/2(\underline{x}-\mu)^{+}C^{-1}(\underline{x}-\mu)} \qquad (17.1)$$

$$\underline{x} = \begin{vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{vmatrix} , \mu = \begin{vmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{vmatrix} , C = \begin{vmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{vmatrix} , |C| = \det C$$

dove μ sia un qualunque vettore di costanti e C una qualunque matrice simmetrica, definita positiva in senso stretto (e quindi non singolare).

Dimostriamo in primo luogo che

$$\mu_X = E\{\underline{x}\} = \mu \tag{17.2}$$
$$C_{XX} = E\{(\underline{x} - \mu)(\underline{x} - \mu)^+\} = C \ . \tag{17.3}$$

A tale scopo notiamo che se

$$\mu = 0$$
$$C = I = \begin{vmatrix} 1 & & 0 & \\ & 1 & & \\ & & \ddots & \\ & 0 & & 1 \end{vmatrix}$$

allora la (17.1) prende forma

$$f(\underline{z}) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\underline{z}^+\underline{z}} = \frac{1}{(2\pi)^{n/2}} e^{-1/2 \sum_{i=1}^{n} z_i^2} \ , \tag{17.4}$$

che è una variabile a n dimensioni le cui componenti sono tutte indipendenti tra loro e tutte normali standardizzate: la v.c. $n-$dimensionale $\underline{Z}$ distribuita secondo la legge (17.4) è detta variabile normale standardizzata a n dimensioni.

Notiamo che per ogni Z_i si ha

$$E\{Z_i\} = 0 \ , \ \sigma^2(Z_i) = 1 \ ;$$

inoltre essendo Z_i indipendente da Z_k

$$c_{ik} = E\{Z_i Z_k\} = 0 \qquad (i \neq k) \ .$$

Ne deduciamo che per la normale standardizzata le (17.2), (17.3) sono valide. Infatti, per quanto appena detto

$$\mu_Z = 0 \tag{17.5}$$

$$C_{ZZ} = \begin{vmatrix} \sigma_1^2 & & 0 & \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{vmatrix} = \begin{vmatrix} 1 & & 0 & \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{vmatrix} = I \tag{17.6}$$

Ora notiamo che (vedi (14.30), (14.31)) essendo C simmetrica e definita positiva, esisterà una matrice K, anch'essa simmetrica e definita positiva, tale che

$$K^2 = C \ : \tag{17.7}$$

dalla (17.7), per note proprietà dei determinanti, segue anche che

$$(\det \ K)^2 = \ \det \ C \ . \tag{17.8}$$

Consideriamo ora la v.c.

$$\underline{\xi} = \mu + K\underline{Z} \ .$$

In base all'Esempio 13.1, essendo $K^+ = K, \xi$ sarà distribuita con densità di probabilità

$$f(\underline{\xi}) = \frac{1}{(2\pi)^{n/2}|K|} e^{-1/2(\underline{\xi} - \mu)^+ (K^2)^{-1} (\underline{\xi} - \mu)} \ ; \tag{17.9}$$

per le (17.7) e (17.8) è chiaro che la (17.9) coincide con la (17.1) e quindi

$$\underline{X} = \underline{\xi} = \mu + K\underline{Z} \ . \tag{17.10}$$

Ora usando il teorema della media si trova (vedi (17.5))

$$\mu_X = \mu + K\mu_Z = \mu$$

che prova la (17.2).

Inoltre, per la legge di propagazione della covarianza, usando la (17.6) si ha

$$C_{XX} = K\ C_{ZZ}K^{+} = K^2 = C$$

che prova la (17.3).

Indicheremo una variabile normale $\underline{X}$ di media μ_X e di covarianza C_{XX} con la formula

$$\underline{X} = N[\mu_X, C_{XX}]\ .$$

Osservazione 17.1: la trasformazione inversa della (17.10) è data da

$$\underline{Z} = K^{-1}(\underline{X} - \mu) \tag{17.11}$$

e viene detta standardizzazione della $\underline{X}$.

Osservazione 17.2: per le variabili normali il concetto di non correlazione e di indipendenza stocastica è equivalente. Infatti supponiamo che X sia una v.c. normale a componenti incorrelate: in questo caso, ricordando che

$$\rho_{ik} = \frac{\sigma_{X_iX_k}}{\sigma_{X_i}\sigma_{X_k}} = \frac{C_{ik}}{\sqrt{C_{ii}C_{kk}}}\ ,$$

si ha

$$C_{ik} = \rho_{ik}\sqrt{C_{ii}C_{kk}} = 0 \qquad (i \neq k)\ .$$

Perciò la matrice di covarianza ha la forma

$$C = \begin{vmatrix} \sigma_1^2 & & 0 & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & 0 & & \sigma_n^2 \end{vmatrix} ,$$

così che risulta

$$(\underline{x} - \mu)^+ C^{-1} (\underline{x} - \mu) = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \ .$$

Si ha inoltre

$$\det C = \sigma_1^2 \sigma_2^2 \ldots \sigma_n^2 \ .$$

Per la (17.1) allora

$$f(\underline{x}) = \frac{1}{(2\pi)^{n/2} \sigma_1 \sigma_2 \ldots \sigma_n} e^{-\frac{1}{2} \sum_i \frac{(x_i - \mu_i)^2}{\sigma_i^2}} =$$
$$= \Pi_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}} = \Pi_i f_{X_i}(x_i) \ :$$

e le X_i risultano stocasticamente indipendenti tra loro.

Osservazione 17.3: nel caso $n = 2$ si ha

$$C = \begin{vmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{vmatrix} = \begin{vmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{vmatrix}$$
$$C^{-1} = \frac{1}{(1-\rho^2)\sigma_X^2\sigma_Y^2} \begin{vmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{vmatrix} =$$
$$= \frac{1}{1-\rho^2} \begin{vmatrix} \frac{1}{\sigma_X^2} & -\frac{\rho}{\sigma_X\sigma_Y} \\ -\frac{\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2} \end{vmatrix}$$

così che la (17.1) prende la consueta forma

$$f(x,y) = \frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_X\sigma_Y} \cdot \qquad (17.12)$$
$$\cdot \quad e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right\}}$$

dove

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

é chiaramente il coefficiente di correlazione tra X e Y.

Notiamo che con la trasformazione

$$\xi = \frac{x-\mu_X}{\sigma_X}$$
$$\eta = \frac{y-\mu_Y}{\sigma_Y}$$

si ottiene una normale doppia le cui componenti sono singolarmente standardizzate e che hanno la distribuzione congiunta

$$f(\xi,\eta) = \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\xi^2 - 2\rho\xi\eta + \eta^2\right\}} \; . \qquad (17.13)$$

Osservazione 17.4: vogliamo trovare la curva di regressione di Y su X e la corrispondente curva di variabilità per una normale doppia. Per definizione

$$\overline{y}(x) = E\{Y|X=x\} \; ;$$

usando le variabili (ξ,η) definite all'Osservazione 17.3, si ha

$$\overline{y}(x) = E\left\{\mu_Y + \sigma_Y\eta|\xi = \frac{x-\mu_X}{\sigma_X}\right\} = \qquad (17.14)$$
$$= \mu_Y + \sigma_Y E\{\eta|\xi\} \; .$$

Ora notiamo che la (17.13) può essere scritta come

$$\begin{aligned} f(\xi,\eta) &= \frac{1}{2\pi(1-\rho^2)^{1/2}} \cdot \\ &\cdot \quad e^{-\frac{1}{2(1-\rho^2)}\left\{(1-\rho^2)\xi^2+\rho^2\xi^2-2\rho\xi\eta+\eta^2\right\}} = \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\xi^2}\frac{1}{\sqrt{2\pi}(1-\rho^2)^{1/2}}e^{-\frac{1}{2}\frac{(\eta-\rho\xi)^2}{1-\rho^2}} : \end{aligned}$$

ricordando che

$$f(\xi,\eta) = f_\xi(\xi) f(\eta|\xi)$$

é evidente che

$$f(\eta|\xi) = \frac{1}{2\pi(1-\rho^2)^{1/2}}e^{-\frac{1}{2}\frac{(\eta-\rho\xi)^2}{1-\rho^2}} ,$$

cioè η, fissato ξ, è distribuita normalmente con media e varianza date da

$$E\{\eta|\xi\} = \rho\xi \qquad \sigma^2(\eta|\xi) = 1-\rho^2 . \tag{17.15}$$

Perciò

$$\overline{\eta}(\xi) = \rho\ \xi = \rho\frac{x-\mu_X}{\sigma_X} ,$$

così tornando alla (17.14) si trova

$$\overline{y}(x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X) . \tag{17.16}$$

Come si vede questa curva di regressione è una retta, detta appunto retta di regressione di Y su X. Quanto alla curva di variabilità, ricordando la seconda delle (17.15) risulta

$$\begin{aligned}\sigma^2(Y|x) &= E\{[Y-\overline{y}(X)]^2|x\} = \\ &= E\{[\mu_Y+\sigma_Y\eta-\mu_Y-\sigma_Y\rho\xi]^2|\xi\} = \\ &= \sigma_Y^2 E\{(\eta-\rho\xi)^2|\xi\} = \\ &= \sigma_Y^2(1-\rho^2)\ :\end{aligned} \tag{17.17}$$

cioè la curva di variabilità è costante.

Notiamo che per $\rho \to \pm 1$, $\sigma^2(Y|X) \to 0$ e la distribuzione tende a concentrarsi sulla retta (17.16), a conferma di quanto abbiamo visto trattando il coefficiente di correlazione lineare.

Osserviamo infine che in modo analogo a questo può essere ricavata la retta di regressione di X su Y che risulta

$$\overline{x}(y) = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\ . \tag{17.18}$$

Entrambe le rette di regressione passano per il punto (μ_X, μ_Y), ma mentre la $\overline{y}(x)$ ha coefficiente angolare $\rho\sigma_Y/\sigma_X$, la $\overline{x}(y)$ ha coefficiente angolare $(1/\rho)(\sigma_Y/\sigma_X)$ (fig. 17.1).

Le due rette possono perciò coincidere se e solo se

$$\rho = \frac{1}{\rho} \to \rho^2 = 1 \to \rho = \pm 1\ ,$$

cioè se la distribuzione è singolare.

Concludiamo il paragrafo riportando, senza dimostrazione, un teorema già precedentemente citato.

Teorema 17.1: tutte le trasformazioni lineari trasformano variabili normali in variabili normali.

$$\left.\begin{array}{l}\underline{X} = N[\mu_X, C_{XX}] \\ \underline{Y} = A\underline{X} + \underline{b}\end{array}\right\} \Rightarrow \underline{Y} = N[A\mu_X + \underline{b},\ AC_{XX}A^+] \tag{17.19}$$

e questo qualsiasi siano le dimensioni n di $\underline{X}$ ed m di $\underline{Y}$.

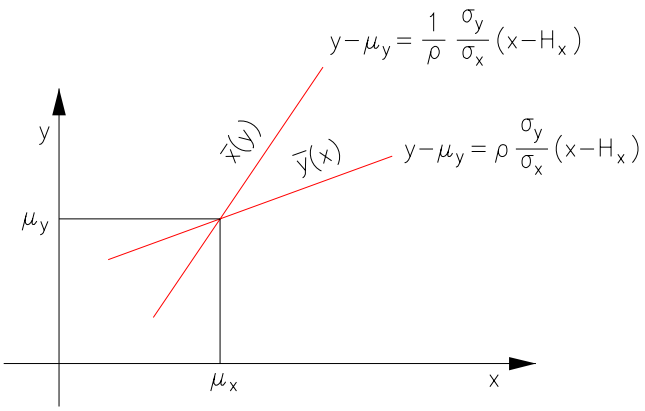


Figura 17.1

Naturalmente se $m > n$, $\underline{Y}$ non potrà essere una variabile regolare poiché sarà distribuita su un sottospazio a n dimensioni di R_m; in tal caso si ha

$$\det\, AC_{XX}A^{+} = 0\ .$$

Se invece $m \leq n$ e la matrice A ha un rango m, allora si può dimostrare che

$$\det\, AC_{XX}A^{+} > 0$$

ed $\underline{Y}$ è una v.c. normale regolare.

Osservazione 17.5: il Teorema 17.1 può essere dimostrato, almeno per il caso $m \leq n$ e rango $(A) = m$, facendo uso delle funzioni generatrici dei momenti.

A tale scopo osserviamo in primo luogo che la $G_{\underline{X}}(\underline{t})$ per una $\underline{X} = N[\mu, C_{XX}]$ è data dalla formula

$$G_{\underline{X}}(\underline{t}) = e^{\underline{t}^{+}\mu + \frac{1}{2}\underline{t}^{+}C_{XX}\underline{t}}\ . \tag{17.20}$$

In effetti la (17.20) è certamente vera per una normale standardizzata; infatti in tal caso, usando l'indipendenza delle componenti di $\underline{Z}$, si ha

$$\begin{aligned} G_{\underline{X}}(\underline{t}) &= E\{e^{\sum t_i Z_i}\} = \Pi\ E\{e^{t_i Z_i}\} = \\ &= \Pi e^{\frac{1}{2}t_i^2} = e^{\frac{1}{2}\sum t_i^2} = e^{\frac{1}{2}\underline{t}^+\underline{t}}\ . \end{aligned} \tag{17.21}$$

che coincide con la (17.20) quando $\mu = 0$ e $C_{XX} = I$.

Per una $\underline{X}$ qualsiasi, usando la rappresentazione (17.10), si trova

$$\begin{aligned} G_{\underline{X}}(\underline{t}) &= E\{e^{\underline{t}^+\underline{X}}\} = E\{e^{\underline{t}^+\mu + \underline{t}^+K\underline{Z}}\} = \\ &= e^{\underline{t}^+\mu}E\{e^{(K\underline{t})^+\underline{Z}}\} = e^{\underline{t}^+\mu}\ e^{1/2\underline{t}^+KK\underline{t}} \end{aligned} \tag{17.22}$$

che appunto coincide con la (17.20).

Ora se vale la (17.19) si ha anche

$$\begin{aligned} G_{\underline{Y}}(\underline{t}) &= E\{e^{\underline{t}^+\underline{Y}}\} = E\{e^{\underline{t}^+\underline{b}} \cdot e^{\underline{t}^+A\underline{X}}\} = \\ &= e^{\underline{t}^+\underline{b}}E\{e^{(A^+\underline{t})^+\underline{X}}\} = e^{\underline{t}^+(\underline{b}+A\mu) + 1/2\underline{t}^+AC_{XX}A^+\underline{t}} \end{aligned}$$

che prova appunto che $Y = N[\underline{b} + A\mu, AC_{XX}A^+]$.

Osservazione 17.6: in base alla (17.11) si ha che la variabile

$$(\underline{X} - \mu)^+C_{XX}^{-1}(\underline{X} - \mu) = \underline{Z}^+\underline{Z} = \sum_{i=1}^{n} Z_i^2 = \chi_n^2$$

risulta χ^2 a n gradi di libertà (vedi Esempio 13.2). Questa osservazione permette facilmente di trovare in R_n una regione simmetrica attorno alla media, μ, in cui sia contenuta una probabilità prefissata p della v.c.

Infatti se $\overline{\chi}^2$ è tale che

$$P(\chi^2 \leq \overline{\chi}^2) = p \qquad (n \text{ gradi di libertà})$$

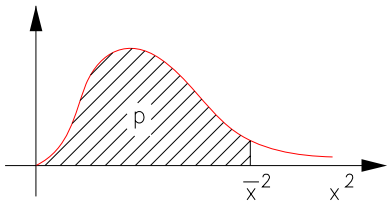


Figura 17.2

allora sarà anche

$$P\left((\underline{X}-\mu)^{+}C_{XX}^{-1}(\underline{X}-\mu)\leq\overline{\chi}^2\right)=p\ :$$

la regione $(\underline{X}-\mu)^{+}C_{XX}^{-1}(\underline{X}-\mu)\leq\overline{\chi}^2$ risulta poi essere un ellissoide e in particolare per $n=2$ essa sarà l'insieme dei punti interni all'ellisse di equazione

$$(\frac{x-\mu_X}{\sigma_X})^2-2\ \rho_{XY}(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y})+(\frac{y-\mu_Y}{\sigma_Y})^2=\overline{\chi}^2\ .$$

Questo ellissoide (o ellisse) viene usualmente chiamato ellissoide (o ellisse) d'errore della v.c. X relativo alla probabilità p; i valori più usati per p sono $p=50\%,\ p=90\%$.

Esempio 17.1: sia (U,V) una normale doppia con matrice di covarianza

$$C_0=\left|\begin{array}{cc}\sigma_U^2 & \sigma_{UV}^2\\ \sigma_{UV} & \sigma_V^2\end{array}\right|\ ;$$

vogliamo sapere se è possibile, dando una rotazione agli assi nel piano (u,v), trovare una nuova normale doppia le cui componenti siano incorrelate.

La trasformazione sarà (fig. 17.3)

$$\left|\begin{array}{c}x\\ y\end{array}\right|=\left|\begin{array}{cccc}\cos\alpha & u+ & \sin\alpha & v\\ -\sin\alpha & u+ & \cos\alpha & v\end{array}\right|=\left|\begin{array}{cc}\cos\alpha & \sin\ \alpha\\ -\sin\ \alpha & \cos\alpha\end{array}\right|\left|\begin{array}{c}u\\ v\end{array}\right|$$

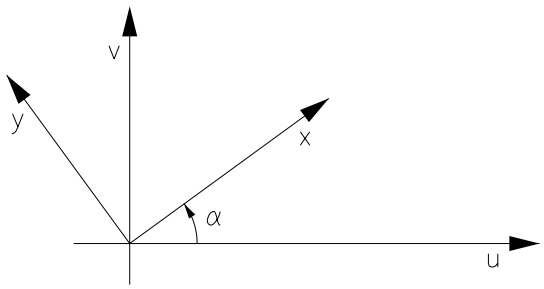


Figura 17.3

Perciò la covarianza di X e Y è la matrice

$$
\begin{aligned}
C_1 &= \begin{vmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{vmatrix} C_0 \begin{vmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{vmatrix} = \\
&= \begin{vmatrix} \sigma_U^2\cos^2\alpha + & -(\sigma_U^2-\sigma_V^2)\cos\alpha\sin\alpha + \\ 2\sigma_{UV}\cos\alpha\sin\alpha + & +\sigma_{UV}(\cos^2\alpha-\sin^2\alpha) \\ +\sigma_V^2\sin^2\alpha & \end{vmatrix} \\
&= \begin{vmatrix} & \\ -(\sigma_U^2-\sigma_V^2)\cos\alpha\sin\alpha + & \sigma_U^2\sin^2\alpha + \\ +\sigma_{UV}(\cos^2\alpha-\sin^2\alpha) & -2\sigma_{UV}\sin\alpha\cos\alpha + \\ & +\sigma_V^2\cos^2\alpha \end{vmatrix}
\end{aligned}
$$

naturalmente risulta $\sigma_{XY} = 0$ se α è scelto in modo tale che

$$\text{tg } 2\alpha = \frac{2\sigma_{UV}}{\sigma_U^2 - \sigma_V^2} \ .$$

Si può provare che anche per una normale n-dimensionale $\underline{X}$, si può trovare una matrice di rotazione R tale che

$$\underline{Y} = R\underline{X}$$

abbia tutte le componenti incorrelate tra loro.

18 Successioni di variabili casuali: convergenza stocastica, teorema di Bernoulli

Sia $\{X_n\}$ una successione di variabili casuali: si dice che $\{X_n\}$ tende stocasticamente a zero per $n \to \infty$ se per ogni $\varepsilon > 0$ prefissato risulta

$$\lim_{n\to\infty} P(|X_n| < \varepsilon) = 1 \ . \tag{18.1}$$

Notiamo che la (18.1) significa sostanzialmente che $\{X_n\}$ tende alla *variabile casuale* X concentrata nell'origine ($P(X = 0) = 1$).

Usando il teorema di Tchebycheff è facile mostrare che vale il seguente lemma.

Lemma 18.1: condizione sufficiente affinché $\{X_n\}$ converga stocasticamente a zero è che

$$\begin{aligned} &\lim_{n\to\infty} E\{X_n\} = 0 \\ &\lim_{n\to\infty} \sigma^2(X_n) = 0 \ . \end{aligned} \tag{18.2}$$

Diremo poi che $\{X_n\}$ converge stocasticamente alla v.c. X se $\{X - X_n\}$ converge stocasticamente a zero.

Questo lemma permette ad esempio di dare una dimostrazione assai semplice del *teorema di Bernoulli*, detto anche legge dei grandi numeri in forma debole.

Teorema 18.1: sia $\{K_n\}$ una successione di variabili binomiali

$$P(k|n) = \binom{n}{k} p^k q^{n-k} \qquad (p+q) = 1$$

e sia $\{\mathbf{f}_n\}$ la successione delle variabili

$$\mathbf{f}_n = \frac{K_n}{n} \ :$$

$\{\mathbf{f}_n\}$ converge stocasticamente a p

$$\lim_{n\to\infty} \mathbf{f}_n = p \ .$$

Infatti si ha

$$E\{\mathbf{f}_n\} = \frac{E\{K_n\}}{n} = \frac{np}{n} = p \text{ (cost.)}$$
$$\sigma^2(\mathbf{f}) = \frac{\sigma^2(K_n)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n} \to 0 \ ;$$

quindi $\mathbf{f}_n - p$ tende stocasticamente a zero.

Osservazione 18.1: l'importanza del teorema di Bernoulli sta nel fatto che fornisce uno schema teorico per l'interpretazione della legge empirica del caso, già citata nei primi paragrafi di queste dispense; in pratica si getta un ponte tra il concetto di frequenza e quello di probabilità. Infatti sia $\mathcal{E}$ un qualsiasi esperimento stocastico i cui risultati sono divisi in due classi: successo e insuccesso. $\mathcal{E}$ sarà descritto da una v.c.

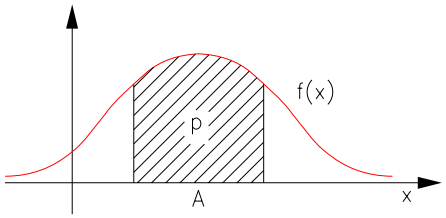
$$P(\xi = 0) = q \quad \text{probabilità di insuccesso}$$
$$P(\xi = 1) = p \quad \text{probabilità di successo.}$$

Si consideri allora una successione di ripetizioni indipendenti di $\mathcal{E}$; $\{K_n\}$ è la successione delle v.c. che descrivono la probabilità di K_{min} successi su n prove.

Perciò

$$\mathbf{f}_n = \frac{K_n}{n}$$

rappresenta la v.c. frequenza relativa di successi su n prove. Se ora prendiamo come $\mathcal{E}$ l'estrazione di una v.c. X qualsiasi e consideriamo come successo l'evento $x \in A$, si avrà ovviamente



$$p = P(x \in A) \ ;$$

Figura 18.1

il teorema perciò afferma che la frequenza relativa di successi, cioè la frequenza di estrazioni da X che ricadono in A, tende stocasticamente al valore p per $n \to \infty$, cioè

$$\lim_{n\to\infty} \frac{n(x_i \in A)}{n} = P(x \in A) \ .$$

Si noti che il modo di tendere delle $\mathbf{f}_n$ a p è stocastico e dunque non è *impossibile* che per una certa successione di estrazioni $\mathbf{f}_n$ si discosti da p, è solo molto improbabile! In questo modo, come si era già detto nel §2, si risolve tramite la definizione assiomatica di probabilità una contraddizione che era invece tipica della definizione di Von Mises.

19 Convergenza in legge: teorema centrale della statistica

Oltre alla convergenza stocastica di una successione di v.c. $\{X_n\}$ ad una v.c. X si può definire anche un altro tipo di convergenza detta appunto in legge: si dice che $\{X_n\}$ tende ad X in legge, $X_n \xrightarrow{\mathcal{L}} X$, se essendo $\{F_n(x)\}$ la successione delle funzioni di distribuzione di $\{X_n\}$ ed $F(x)$ la funzione di distribuzione di X risulta

$$\lim_{n\to\infty} F_n(x) = F(x) \ , \tag{19.1}$$

quasi ovunque in x, ovvero in tutti punti di continuità della F. La convergenza in legge di X_n ad X può essere valutata anche considerando le funzioni caratteristiche $C_{X_n}(t), C_X(t)$ o, quando esistano, le funzioni generatrici dei momenti, $G_{X_n}(t), G_X(t)$.

Vale infatti il seguente

Teorema 19.1: sia X_n una successione di v.c. tali che

$$C_{X_n}(t) \to C_X(t) \ , \forall t$$

dove $C_X(t)$ risulti una funzione continua in $t = 0$, allora $C_X(t)$ é una funzione caratteristica di una v.c. X ed $X_n \xrightarrow{\mathcal{L}} X$. Sia X_n una successione di v.c. con funzioni generatrici $G_{X_n}(t)$ e supponiamo che

$$G_{X_n}(t) \to G_X(t)$$

$\forall t$ in un intorno fissato dell'origine, $G_X(t)$ essendo la funzione generatrice di una v.c. X, allora $X_n \xrightarrow{\mathcal{L}} X$.

La convergenza in legge è particolarmente importante nello studio del comportamento asintotico di somme di variabili casuali:

$$S_n = \sum_{i=1}^{n} X_i \tag{19.2}$$

Si può infatti dimostrare che sotto opportune ipotesi sulla successione delle $\{X_i\}$ la successione $\{S_n\}$ tende asintoticamente in legge ad una distribuzione normale: si tratta del celebre teorema centrale della statistica di cui sono state date molte enunciazioni. Noi ci limiteremo qui alla più semplice che copre però alcuni importanti casi.

Teorema 19.2: sia $\{X_i\}$ una successione di variabili indipendenti, tutte con la stessa distribuzione e con

$$E\{X_i\} = \mu$$
$$\sigma^2(X_i) = \sigma^2$$

allora la successione

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

ha l'andamento asintotico in legge

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

qualunque sia la distribuzione inziale delle $\{X_i\}$, ovvero

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{L}} N[0,1] \tag{19.3}$$

Accenneremo alla dimostrazione della (19.3).

Per dimostrare la (19.3) basterà provare che, definita la successione

$$R_n = \frac{S_n - n\mu}{\sqrt{n}} = \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sqrt{n}} \ , \tag{19.4}$$

si ha

$$\lim_{n\to\infty} g_{R_n}(t) = e^{\frac{1}{2}\sigma^2 t} \ , \tag{19.5}$$

così che

$$\lim_{n\to\infty} R_n = \sigma Z \quad \text{(in legge)} \ , \tag{19.6}$$

ovvero

$$S_n - n\mu \sim \sqrt{n}\ \sigma Z = N[0, n\sigma^2] \ . \tag{19.7}$$

A tale scopo basta notare che le variabili indipendenti ed identicamente distribuite

$$Y_i = \frac{X_i - \mu}{\sqrt{n}}$$

hanno tutte la stessa funzione generatrice

$$g_Y(t) = 1 + \frac{1}{2}\frac{\sigma_X^2}{n} t^2 + O\left(\frac{1}{n}\right) \ ,$$

in quanto vale

$$E\{Y_i\} = 0$$
$$\sigma^2\{Y_i\} = \frac{\sigma_X^2}{n} \ .$$

Pertanto, sfruttando l'indipendenza delle Y_i,

$$\begin{aligned} g_{R_n}(t) &= E\{e^{t\sum_i Y_i}\} = \Pi_i \ E\{e^{tY_i}\} = [g_Y(t)]^n = \\ &= \left[1 + \frac{1}{2}\frac{\sigma_X^2}{n}t^2 + O\left(\frac{1}{n}\right)\right]^n \ . \end{aligned} \tag{19.8}$$

Dalla (19.8), prendendo il limite per $n \to \infty$ e ricordando la definizione del numero e, discende la (19.5), cioè la tesi che volevamo dimostrare.

Senza ulteriori dimostrazioni generalizziamo un poco il Teorema 19.2 considerando una successione $\{X_i\}$ di v.c. indipendenti ma non più necessariamente identiche; in questo caso possiamo ad esempio rifarci ad una formulazione, fra le molte, dovuta a Lyapunov

Teorema 19.3: sia $\{X_i\}$ una successione di v.c. indipendenti e tali che, per un qualche $\delta > 0$,

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E\{|X_i - \mu_i|^{2+\delta}\}}{[\sum_{i=1}^n \sigma_i^2]^{1+\frac{\delta}{2}}} = 0 \ , \tag{19.9}$$

allora $\dfrac{S_n - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$ tende in legge ad una Z.
Notiamo solo che se le X_i sono anche tutte identiche tra loro, vale la relazione

$$\frac{\sum_{i=1}^n E\{|X_i - \mu_i|^{2+\delta}\}}{[\sum_{i=1}^n \sigma_i^2]^{1+\frac{\delta}{2}}} = \frac{n\overline{\mu}_{2+\delta}}{n^{(1+\frac{\delta}{2})}\sigma^2}$$

che tende a zero per ogni $\delta > 0$, a condizione che $\overline{\mu}_{2+\delta} < +\infty$.

Osservazione 19.1: notiamo che nell'applicare i Teoremi 19.1 e 19.2 non è richiesto che le $\{X_i\}$ siano v.c. continue.

Così ad esempio se le X_i sono variabili elementari del tipo

$$X_i \begin{cases} -1 & 1 \\ 1/2 & 1/2 \end{cases},$$

é ancora vero che $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \sim N[0,1]$, risultato questo noto anche sotto il nome di teorema di Laplace-De Moivre.

Osservazione 19.2: una importante applicazione di questo teorema è nella interpretazione che tramite esso si può dare al fatto sperimentalmente riconosciuto che gli errori di misura tendono a distribuirsi normalmente: ciò almeno quando il procedimento di misura sia usato vicino al limite della sua precisione massima (é ad esempio evidente che misurando l'altezza di un uomo con l'approssimazione del metro non si darà mai luogo ad una distribuzione normale degli errori!).

L'interpretazione di cui stiamo parlando è la seguente: gli errori di misura dipendono da una serie di fattori ambientali, strumentali e soggettivi che hanno, preso ciascuno isolatamente, una influenza impercettibile sul procedimento di misura; ciascuno di questi fattori perciò assume l'aspetto di una v.c. che per lo più può essere ritenuta indipendente dalle altre.

Fattori di questo tipo possono essere il grado di umidità dell'aria, la sua temperatura, la luminosità del giorno oppure la non perfetta planeità di uno specchio, la presenza di microcorrenti indotte (rumore) in un circuito oppure il grado di affaticamento dell'occhio di un operatore, ecc. Come si é detto, ogni singolo fattore ha una influenza trascurabile sul procedimento di misura, ma tutti messi insieme producono un effetto sensibile, l'errore di misura, che sarà descritto da una variabile casuale somma di molte altre, come in (19.2): per il teorema centrale tale errore di misura tenderà perciò ad essere distribuito normalmente.

Osservazione 19.3: il teorema centrale permette spesso di usare la normale come distribuzione approssimata per quantità statisticamente importanti. Valga come esempio il seguente; sia X una v.c. comunque distribuita e sia $\{X_1, X_2 \ldots X_n\}$ la v.c. n-dimensionale generata pensando di ripetere n estrazioni indipendenti da X; questa variabile descrive evidentemente i campioni di numerosità n della X.

Un indice di grande importanza nella descrizione del campione è la media campionaria

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i \ ; \tag{19.10}$$

notiamo che si può anche scrivere

$$M = \sum_{i=1}^{n} (\frac{X_i}{n}) \ .$$

Ora le X_i/n sono variabili per ipotesi tutte indipendenti e se era

$$\begin{aligned} E\{X_i\} &= \mu \\ \sigma^2(X_i) &= \sigma^2 \end{aligned}$$

sarà anche

$$\begin{aligned} E\left\{\frac{X_i}{n}\right\} &= \frac{\mu}{n} \\ \sigma^2\left(\frac{X_i}{n}\right) &= \frac{\sigma^2}{n^2} \ . \end{aligned}$$

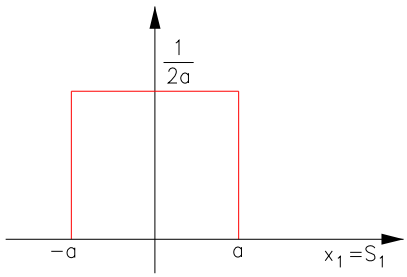
Se supponiamo che il campione sia numeroso (n grande) potremo applicare ad M il teorema centrale e dire che, qualsiasi fosse la distribuzione iniziale di X, M sarà

$$M \sim N\left[n\frac{\mu}{n}, n\frac{\sigma^2}{n^2}\right] = N\left[\mu, \frac{\sigma^2}{n}\right] \ . \tag{19.11}$$

Si noti che se si volesse ricavare la distribuzione esatta di M, ovvero $\sum_i X_i$, si dovrebbero calcolare n integrali di convoluzione della $\mathbf{f}_X(x)$ con se stessa, cosa questa non sempre agevole e sicuramente più complicata dell'uso della (19.11).

Esempio 19.1: per esemplificare il modo in cui una somma di variabili tende alla distribuzione normale, consideriamo la somma di tre variabili indipendenti uniformemente distribuite sull'intervallo $[-a, a]$

$$f_{X_1}(x_1) = \frac{1}{2a} \ ; \ (-a \leq x_1 \leq a)$$


1
2a
−a
a
x1 =S1

$$S_2 = X_1 + X_2$$

$$f_{S_2}(s_2) = \begin{cases} \dfrac{s_2 + 2a}{4a^2} & ; \quad (-2a \le s_2 \le 0) \\ \dfrac{2a - s_2}{4a^2} & ; \quad (0 \le s_2 \le 2a) \end{cases}$$

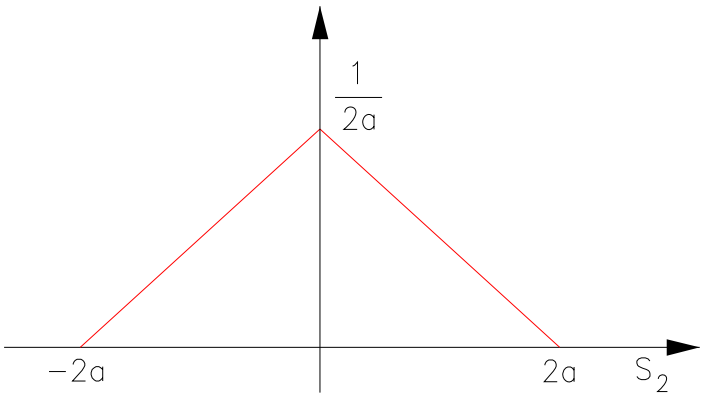


Figura 19.2

$$S_3 = X_1 + X_2 + X_3$$

$$f_{S_3}(s_3) = \begin{cases} \dfrac{(s_3 + 3a)^2}{16a^3} & ; \quad (-3a \le s_3 \le -a) \\ \dfrac{3}{8a} - \dfrac{s_3^2}{8a^3} & ; \quad (-a \le s_3 \le a) \\ \dfrac{(s_3 - 3a)^2}{16a^3} & ; \quad (a \le s_3 \le 3a) \end{cases}$$

Come si vede, dopo solo tre addendi la distribuzione ha già assunto un andamento a campana tipico della normale. Per un confronto più preciso si consideri che per $a = 1$ si ha

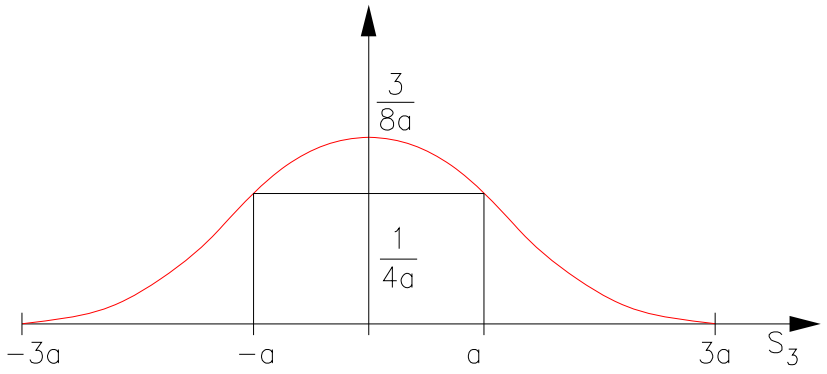


Figura 19.3

$$\sigma^2(X_i) = 1/3$$
$$\sigma^2(S_3) = 1 \ ,$$

così che $f_{S_3}(s_3)$ va confrontata con $Z = N[0,1]$: si può allora fare la seguente tabella comparativa (approssimata alla seconda cifra)

Z	$f_Z(z)$	$f_{s_3}(z)$
0	0,40	0,38
1	0,24	0,25
2	0,05	0,06
3	0,00	0,00

che mostra già un buon adattamento di S_3 a $N[0,1]$.