

ANALISI MATEMATICA

Ottavio Caligaris - Pietro Oliva

ELEMENTI DI PROBABILITÀ E STATISTICA.

Il calcolo delle probabilità nasce agli inizi del '700 per rispondere alle istanze prodotte dal gioco d'azzardo.

Il suo scopo era inizialmente introdurre una stima numerica dei rischi ad esso connessi e successivamente lo stesso metodo fu esteso per studiare i fenomeni caratterizzati da elementi di incertezza.

Vediamo innanzi tutto di stabilire quali sono gli elementi necessari per parlare di probabilità

**Occorre innanzi tutto considerare una famiglia di eventi che costituiscono lo spazio su cui definiremo la probabilità.
Possiamo ad esempio associare all'idea di evento quella di sottoinsieme di un insieme dato che chiameremo evento certo.**

1. Spazi di probabilità

Consideriamo, ad esempio, il più semplice tra i giochi d'azzardo, e cioè il lancio di una moneta, possiamo dire che gli eventi possibili sono

- L'uscita di Testa $\{\mathbf{T}\}$
- L'uscita di Croce $\{\mathbf{C}\}$

mentre l'insieme $\mathcal{U} = \{\mathbf{T}, \mathbf{C}\}$ costituisce l'evento certo.

In questo caso è naturale definire la probabilità degli eventi che entrano in gioco:

$$(28.1) \quad \mathcal{P}(T) = \frac{1}{2}$$

$$(28.2) \quad \mathcal{P}(C) = \frac{1}{2}$$

$$(28.3) \quad \mathcal{P}(\mathcal{U}) = \mathcal{P}(T) + \mathcal{P}(C) = 1$$

$$(28.4) \quad \mathcal{P}(\emptyset) = 0$$

giustificando la definizione con il fatto che su due possibili uscite una sola è favorevole nel caso si consideri T o C , mentre entrambe vanno bene nel caso si consideri l'unione di T e C . È ovvio che con ciò supponiamo che T e C si presentino con ugual frequenza, cioè che la moneta sia non truccata.

Un secondo esempio di spazio di probabilità si può costruire considerando il caso del lancio di due dadi.

Se le facce sono numerate, come al solito, da 1 a 6 possiamo identificare l'esito del lancio con la coppia di numeri (i, j) (punteggio) che si leggono sulla faccia superiore del primo e del secondo dado.

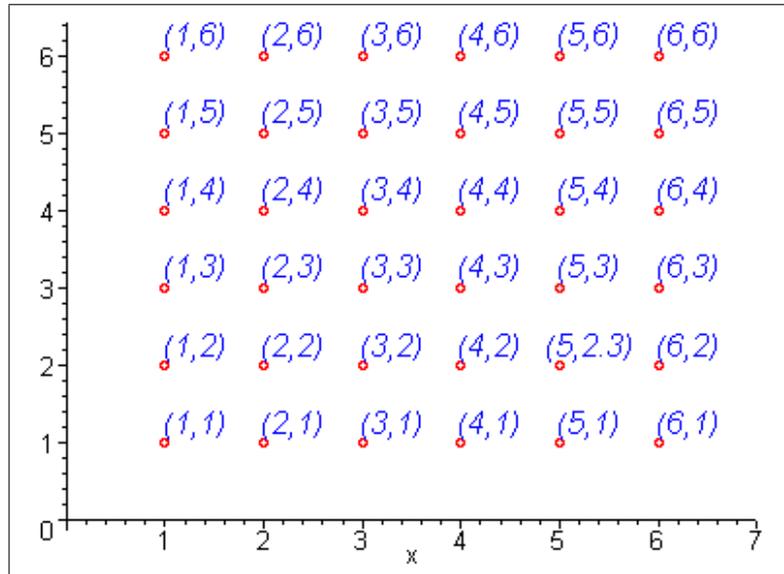


FIGURA 28.1. Lo spazio \mathcal{U} degli eventi nel caso del lancio di due dadi

In tal modo possiamo identificare ciascuna delle 36 possibili uscite (eventi) con il punto del piano cartesiano di coordinate (i, j) ; indicheremo tale evento con il simbolo $A_{i,j}$, (si veda la figura 28.1).

Poichè nel caso di dadi non truccati ogni evento è equiprobabile possiamo affermare che la probabilità di $A_{i,j}$ è data da

$$\mathcal{P}(A_{i,j}) = \frac{1}{36}$$

Ovviamente possiamo considerare anche altri eventi, ad esempio possiamo cercare di stimare la probabilità che si presenti l'evento $A_{1,4}$ oppure l'evento $A_{5,6}$; possiamo indicare il nuovo evento come

$$B = A_{1,4} \cup A_{5,6} = \{A_{1,4}, A_{5,6}\}$$

ed è ragionevole stimare che, poichè accettiamo 2 eventi su 36 possibili,

$$\mathcal{P}(B) = \frac{2}{36} = \frac{1}{36} + \frac{1}{36}$$

Se vogliamo stimare la probabilità che si presenti uno qualunque degli eventi $A_{i,j}$, cioè se vogliamo stimare la probabilità che si presenti l'evento

$$\mathcal{U} = \{A_{i,j} : i, j = 1..6\}$$

poichè accettiamo 36 possibilità su 36 possiamo dire che

$$P(\mathcal{U}) = 1$$

e che \mathcal{U} è l'evento certo.

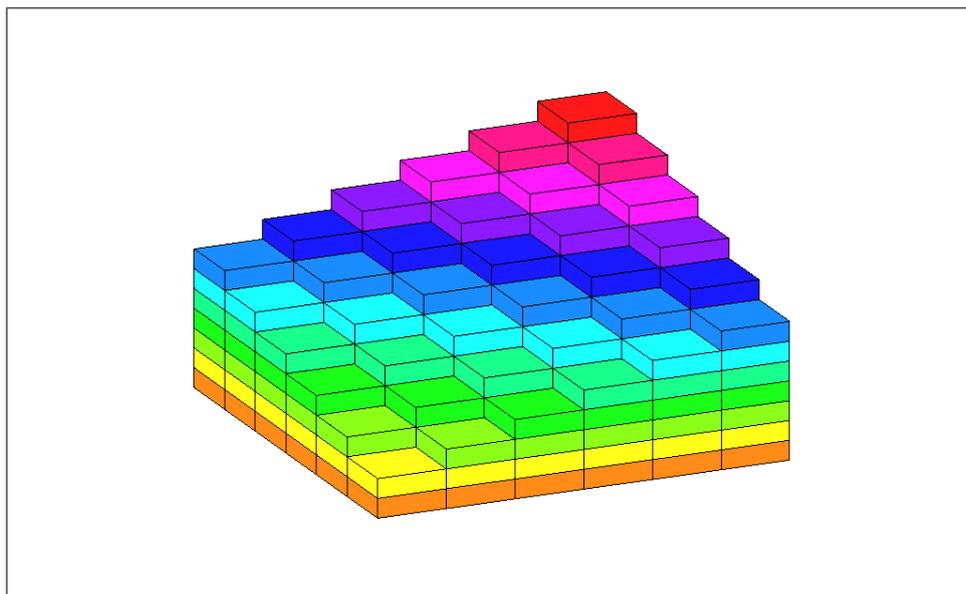


FIGURA 28.2. Istogramma relativo al lancio di due dadi, (supposti distinti)

In generale possiamo parlare di spazio di probabilità finito se è assegnata una famiglia di eventi

$$\mathcal{F} = \{A_i : i = 1..N\}$$

soddisfacente le seguenti condizioni

- $\emptyset \in \mathcal{F}$
- $A_1, A_2 \in \mathcal{F} \implies A_1 \cup A_2 \in \mathcal{F}$

ed è assegnata una funzione

$$\mathcal{P} : \mathcal{F} \rightarrow \mathbb{R}$$

che associa ad ogni $A \in \mathcal{F}$ un valore reale $\mathcal{P}(A)$, che chiamiamo probabilità che l'evento accada, soddisfacente le seguenti proprietà:

- $\mathcal{P}(A) \geq 0$
- **se $\mathcal{U} = \bigcup_{A \in \mathcal{F}} A \in \mathcal{F}$, si ha $\mathcal{P}(\mathcal{U}) = 1$**
- **se $A_1, A_2 \in \mathcal{F}$ sono tali che $A_1 \cap A_2 = \emptyset$ allora**

$$\mathcal{P}(A_1 \cup A_2) = \mathcal{P}(A_1) + \mathcal{P}(A_2)$$

Osserviamo che, dal momento che la famiglia dei possibili eventi è finita, avremo che

$$\mathcal{U} = \bigcup_{A \in \mathcal{F}} A$$

\mathcal{U} rappresenta l'evento certo ed anche l'ambiente in cui si individuano gli eventi.

Per assegnare uno spazio di probabilità, quindi, è necessario assegnare un insieme \mathcal{U} una famiglia \mathcal{F} di sottoinsiemi di \mathcal{U} ed una funzione P definita su \mathcal{F} a valori in \mathbb{R} .

Ci riferiremo quindi ad uno spazio di probabilità come ad un terna $(\mathcal{U}, \mathcal{F}, \mathcal{P})$

Nel caso dell'esempio precedente la famiglia \mathcal{F} è costituita da ciascuno degli eventi (i, j) che abbiamo rappresentato nel piano con i punti $A_{i,j}$ e dalle unioni finite di essi.

Ad esempio

$$B = \{A_{1,3}, A_{4,1}\} = A_{1,3} \cup A_{4,1}$$

è l'evento che accade se almeno una tra le uscite $A_{1,3}$ e $A_{4,1}$ accade.

Poichè le uscite del lancio dei dadi sono ritenute equiprobabili possiamo affermare che $\mathcal{P}(A_{i,j}) = \frac{1}{36}$ per ogni possibile uscita $A_{i,j}$ e quindi, ad esempio,

$$\mathcal{P}(B) = \mathcal{P}(A_{1,3} \cup A_{4,1}) = \mathcal{P}(A_{1,3}) + \mathcal{P}(A_{4,1}) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$$

Ovviamente

$$\mathcal{P}(U) = \mathcal{P}\left(\bigcup_{A \in \mathcal{F}} A\right) = 36 \frac{1}{36} = 1$$

Come abbiamo già osservato, ci riferiamo ad uno spazio di probabilità come ad una terna $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ in cui \mathcal{U} è lo spazio di probabilità, \mathcal{F} è la famiglia di tutti gli eventi e \mathcal{P} è la misura di probabilità definita su \mathcal{U} .

Si possono provare i seguenti risultati:

- $0 \leq \mathcal{P}(A) \leq 1$ per ogni $A \in \mathcal{F}$
- $\mathcal{P}(\emptyset) = 0$
- Se $A_1, A_2 \in \mathcal{F}$ $A_1 \subset A_2$ allora

$$\mathcal{P}(A_1) \subset \mathcal{P}(A_2)$$

ed inoltre

$$\mathcal{P}(A_2 \setminus A_1) = \mathcal{P}(A_2) - \mathcal{P}(A_1)$$

- $\mathcal{P}(A^c) = 1 - \mathcal{P}(A)$ per ogni $A \in \mathcal{F}$
- $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$
- Se $B = \bigcup_{i=1}^N A_i$ e se $A_i \cap A_j = \emptyset$ allora

$$\mathcal{P}(B) = \sum_{i=1}^N \mathcal{P}(A_i)$$

- $\mathcal{P}(A) = \mathcal{P}(A \cap B) + \mathcal{P}(A \cap B^c)$
- se $A_i \cap A_j = \emptyset$, ed esiste i_0 tale che $A \subset A_{j_0}$, allora

$$(28.5) \quad \boxed{\mathcal{P}(A) = \mathcal{P}(A \cap A_1) + \mathcal{P}(A \cap A_2) + \dots + \mathcal{P}(A \cap A_N)}$$

DEFINIZIONE 28.1. Se $A, B \in \mathcal{F}$ definiamo probabilità di A condizionata a B e la denotiamo con $\mathcal{P}(A|B)$ il valore

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

La probabilità di A condizionata a B , $\mathcal{P}(A|B)$, definisce la probabilità di accadimento di A nel caso sia accaduto B

Naturalmente si ha

$$\mathcal{P}(A \cap B) = \mathcal{P}(A|B)\mathcal{P}(B)$$

Nel caso in cui

$$\mathcal{P}(A|B) = \mathcal{P}(A)$$

diciamo che A e B sono eventi indipendenti, (la probabilità di accadimento di A non è cambiata dal fatto che B è accaduto).

In tal caso si ha

$$\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)$$

Si può dimostrare che

Se $A_i \cap A_j = \emptyset$, ed esiste i_0 tale che $A \subset A_{j_0}$, allora

$$\mathcal{P}(A) = \mathcal{P}(A_1)\mathcal{P}(A|A_1) + \mathcal{P}(A_2)\mathcal{P}(A|A_2) + \dots + \mathcal{P}(A_N)\mathcal{P}(A|A_N)$$

e ne segue il seguente teorema di Bayes

TEOREMA 28.1. - di Bayes - Se $A_1, A_2, \dots, A_N \in \mathcal{F}$ sono eventi tali che $A_i \cap A_j = \emptyset$ e $\bigcup_{i=1}^N A_i$ allora

$$\mathcal{P}(A_k|B) = \frac{\mathcal{P}(A_k)\mathcal{P}(A|A_k)}{\sum_{i=1}^N \mathcal{P}(A_i)\mathcal{P}(A|A_i)}$$

Per capire perchè il teorema sia vero, consideriamo il caso in cui $N = 2$; allora

- $A_1 \cup A_2 \supset A$
- $A_1 \cap A_2 = \emptyset$
- $\mathcal{P}(B) = \mathcal{P}(B \cap (A_1 \cup A_2)) = \mathcal{P}(B \cap A_1) + \mathcal{P}(B \cap A_2)$

e

$$\mathcal{P}(A_1|B) = \frac{\mathcal{P}(A_1 \cap B)}{\mathcal{P}(B)}$$

$$\mathcal{P}(B|A_1) = \frac{\mathcal{P}(B \cap A_1)}{\mathcal{P}(A_1)}$$

da cui

$$\mathcal{P}(A_1|B)\mathcal{P}(B) = \mathcal{P}(B \cap A_1) = \mathcal{P}(B|A_1)\mathcal{P}(A_1)$$

Ne segue che

$$\mathcal{P}(A_1|B) = \frac{\mathcal{P}(B \cap A_1)}{\mathcal{P}(B)} = \frac{\mathcal{P}(B \cap A_1)\mathcal{P}(A_1)}{\mathcal{P}(B \cap A_1)\mathcal{P}(A_1) + \mathcal{P}(B \cap A_2)\mathcal{P}(A_2)}$$

2. Qualche applicazione della regola di Bayes.

2.1. Si consideri una popolazione in cui è diffusa una malattia e si supponga di voler sottoporre l'intera popolazione ad un test con lo scopo di determinare per ciascuna persona se è affetta o no dalla malattia.

Una situazione di questo genere si presenterebbe se si considerasse l'opportunità di procedere ad uno screening di massa per rivelare, ad esempio, la diffusione dell'AIDS.

In tal caso sarebbe ragionevole supporre che:

- Nella popolazione 1 persona su 1000 è affetta dalla malattia (**incidenza della malattia**).
- Il test usato per rivelare la malattia è positivo (e quindi indica la presenza della malattia) in 95 casi per ogni 100 persone malate esaminate (**sensitività del test**).
- Il test usato per rivelare la malattia è positivo (e quindi indica la presenza della malattia) in 1 caso per ogni 100 persone sane esaminate (**specificità del test**).

Indichiamo con

- A l'evento "infetto"
- A' l'evento "sano"
- T_p l'evento "test positivo"
- T_n l'evento "test negativo"

Indichiamo inoltre con

- a l'incidenza della malattia (nel caso prima citato $a = 1/1000 = 0.001$)
- p la sensitività del test (nel caso prima citato $p = 95/100 = 0.95$)
- q la specificità del test (nel caso prima citato $q = 1/100 = 0.01$)

Possiamo facilmente calcolare che

$$\begin{aligned} \mathcal{P}(A) &= a & \mathcal{P}(A') &= 1 - a \\ \mathcal{P}(T_p|A) &= p & \mathcal{P}(T_n|A) &= 1 - p \\ \mathcal{P}(T_p|A') &= q & \mathcal{P}(T_n|A') &= 1 - q \end{aligned}$$

da cui

$$\begin{aligned} \mathcal{P}(T_p) &= \mathcal{P}(T_p \cap A) + \mathcal{P}(T_p \cap A') = \\ &= \mathcal{P}(T_p|A)\mathcal{P}(A) + \mathcal{P}(T_p|A')\mathcal{P}(A') = pa + q(1 - a) \end{aligned}$$

Pertanto, applicando il teorema di Bayes

$$\mathcal{P}(A|T_p) = \frac{\mathcal{P}(T_p|A)\mathcal{P}(A)}{\mathcal{P}(T_p|A)\mathcal{P}(A) + \mathcal{P}(T_p|A')\mathcal{P}(A')} = \frac{pa}{pa + q(1 - a)}$$

Se utilizziamo i valori di incidenza sensitività e specificità prima introdotti, possiamo ricavare che:

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.0874$$

e questo significa che la probabilità che una persona in cui il test ha dato esito positivo sia effettivamente infetto è inferiore al 9%.

Se ne deduce la poca convenienza ad effettuare uno screening di massa.

Vale la pena osservare che neppure test più precisi consentono di raggiungere risultati più significativi.

Infatti se supponiamo di migliorare il test in modo da avere la certezza di individuare la malattia nelle persone infette, cioè se supponiamo la sensitività $p = 100/100 = 1$, mantenendo $a = 0.001$ e $q = 0.01$, otteniamo

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.091$$

In realtà aumentando la sensitività, la specificità necessariamente diminuisce; se supponiamo che

$$a = 0.001 \quad , \quad p = 1 \quad , \quad q = 0.002$$

la situazione peggiora e si ha addirittura

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.05$$

Solo migliorando la specificità senza peggiorare la sensitività si ottiene qualcosa di meglio; infatti se

$$a = 0.001 \quad , \quad p = 0.95 \quad , \quad q = 1/300 = 0.0034$$

si ha

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.22$$

Un miglioramento decisivo nell'affidabilità dei risultati si ottiene invece nel caso in cui l'incidenza della malattia sia alta; infatti per

$$a = 0.01 \quad , \quad p = 0.95 \quad , \quad q = 0.01$$

si ha

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.5$$

e addirittura

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.9$$

nel caso in cui l'incidenza salga ad $a = .1$.

2.2. Un secondo interessante esempio si ha considerando la seguente situazione.

Si supponga che un capo di governo debba prendere una decisione di politica economica e disponga di tre consiglieri che identifichiamo con

$A \quad B \quad C$

Si supponga di avere una stima della loro affidabilità descritta come segue:

- la probabilità che il consigliere A abbia un'opinione corretta è $1/6$,
- la probabilità che il consigliere B abbia un'opinione corretta è $1/3$,
- la probabilità che il consigliere C abbia un'opinione corretta è $1/2$.

Indichiamo con

A_r l'evento "il consigliere A ha espresso un'opinione corretta",

B_r l'evento "il consigliere B ha espresso un'opinione corretta",

C_r l'evento "il consigliere C ha espresso un'opinione corretta".

Esprimiamo l'affidabilità dei consiglieri assegnando agli eventi A_r, B_r e C_r opportuni valori di probabilità

$$\mathcal{P}(A_r) = 1/6 \quad \mathcal{P}(B_r) = 1/3 \quad \mathcal{P}(C_r) = 1/2$$

Il capo di governo chiede ai suoi consiglieri una previsione sulle conseguenze che la sua decisione avrà sul tasso di disoccupazione, a distanza di un anno.

Al quesito i tre consiglieri rispondono secondo i valori descritti nella tabella che segue, dove nella colonna D sono riportate le probabilità che la disoccupazione diminuisca, nella colonna S che rimanga stabile e nella colonna I che il tasso di disoccupazione aumenti.

	D	S	I
A	1/10	1/10	8/10
B	6/10	2/10	2/10
C	2/10	6/10	2/10

TABELLA 28.1. Previsioni dei consiglieri sul tasso di disoccupazione.

Trascorso un anno si rileva che il tasso di disoccupazione è aumentato; alla luce di questo fatto le percentuali di affidabilità dei tre consiglieri vanno riviste e ciò può essere fatto ridefinendo la probabilità di A_r , B_r e C_r alla luce del fatto che l'evento I si è verificato.

Dovremo, in altre parole valutare

$$\mathcal{P}(A_r|I) = \frac{\mathcal{P}(I|A_r)\mathcal{P}(A_r)}{\mathcal{P}(I)}$$

$$\mathcal{P}(B_r|I) = \frac{\mathcal{P}(I|B_r)\mathcal{P}(B_r)}{\mathcal{P}(I)}$$

$$\mathcal{P}(C_r|I) = \frac{\mathcal{P}(I|C_r)\mathcal{P}(C_r)}{\mathcal{P}(I)}$$

ma

$$\begin{aligned} \mathcal{P}(I) &= \mathcal{P}(I \cap A_r) + \mathcal{P}(I \cap B_r) + \mathcal{P}(I \cap C_r) = \\ &= \mathcal{P}(I|A_r)\mathcal{P}(A_r) + \mathcal{P}(I|B_r)\mathcal{P}(B_r) + \mathcal{P}(I|C_r)\mathcal{P}(C_r) \end{aligned}$$

Ora, la probabilità che avvenga I ammesso che un consigliere abbia ragione si legge nell'ultima colonna della tabella, per cui

$$\mathcal{P}(I|A_r) = 8/10$$

$$\mathcal{P}(I|B_r) = 2/10$$

$$\mathcal{P}(I|C_r) = 2/10$$

per cui

$$P(I) = \frac{8}{10} \frac{1}{6} + \frac{2}{10} \frac{1}{3} + \frac{2}{10} \frac{1}{2} = \frac{3}{10}$$

Ne viene che le nuove percentuali di affidabilità sono

$$\mathcal{P}(A_r|I) = \frac{\mathcal{P}(I|A_r)\mathcal{P}(A_r)}{\mathcal{P}(I)} = \frac{\frac{8}{10} \frac{1}{6}}{\frac{3}{10}} = \frac{4}{9}$$

$$\mathcal{P}(B_r|I) = \frac{\mathcal{P}(I|B_r)\mathcal{P}(B_r)}{\mathcal{P}(I)} = \frac{\frac{2}{10} \frac{1}{3}}{\frac{3}{10}} = \frac{2}{9}$$

$$\mathcal{P}(C_r|I) = \frac{\mathcal{P}(I|C_r)\mathcal{P}(C_r)}{\mathcal{P}(I)} = \frac{\frac{2}{10} \frac{1}{2}}{\frac{3}{10}} = \frac{1}{3}$$

3. Qualche richiamo di calcolo combinatorio.

Per studiare un po' di probabilità discreta è utile conoscere qualche elemento di calcolo combinatorio.

Il calcolo combinatorio si occupa di stabilire il numero delle possibili uscite di semplici esperimenti; si fonda essenzialmente sul principio seguente:

Se un esperimento ha n_1 possibili esiti, un secondo esperimento ha n_2 possibili esiti, un terzo esperimento ha n_3 possibili esiti, allora il numero dei possibili esiti della sequenza dei tre esperimenti è

$$n_1 n_2 n_3$$

Le più comuni conseguenze di questo principio portano a un certo numero di definizioni che descriviamo brevemente.

3.1. Disposizioni di n elementi a k a k . Parliamo di **disposizioni** (o anche, se $k = n$, di **permutazioni**) di n elementi a k a k quando consideriamo i gruppi di k elementi scelti tra gli n dati.

Riteniamo due gruppi distinti se differiscono per un elemento o per l'ordine con cui gli elementi sono scelti.

Indichiamo con

$${}_n D_k$$

il numero delle disposizioni di n elementi a k a k .

Poichè per il primo elemento di ciascun gruppo abbiamo n scelte, per il secondo ne abbiamo $(n - 1)$ per il terzo ne abbiamo $(n - 2)$ e così via, possiamo calcolare che

$${}_n D_k = n(n - 1)(n - 2)(n - 3)\dots(n - (k - 1)) = \frac{n!}{(n - k)!}$$

Il numero delle disposizioni di n elementi ad n ad n , cioè delle permutazioni, risulta

$$P_n = {}_n D_n = n!$$

Qualora gli n elementi da cui si sceglie presentino sottogruppi di elementi uguali, siano $n_1, n_2, n_3, \dots, n_k$ le permutazioni risultano in numero di

$$\frac{n!}{n_1! n_2! n_3! \dots n_k!}$$

3.2. Combinazioni di n elementi a k a k . Parliamo di **combinazioni** di n elementi a k a k quando consideriamo gruppi di k elementi scelti tra gli n dati senza distinguerli in base all'ordine degli elementi.

Indichiamo con

$${}_n C_k$$

il numero delle combinazioni di n elementi a k a k .

Poichè un gruppo di k elementi ammette $k!$ diversi modi possiamo calcolare che

$${}_n C_k = \frac{{}_n D_k}{k!} = \frac{n(n-1)(n-2)(n-3)\dots(n-(k-1))}{k!} = \frac{n!}{k!(n-k)!}$$

Il numero ${}_n C_k$ si chiama coefficiente binomiale e si indica con

$${}_n C_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Ricordiamo che i coefficienti binomiali possono essere ricavati dal triangolo di Tartaglia e che trovano una importante applicazione nella formula del binomio di Newton.

$$(a+b)^n = \sum_0^n \binom{n}{k} a^k b^{n-k}$$

3.3. Campioni ordinati. È anche utile ricordare qualche formula per stimare il numero di possibili campioni estratti da una popolazione.

Per aiutarci assimiliamo la popolazione ad un un'urna piena di palline e l'estrazione degli elementi del campione all'estrazione delle palline dall'urna.

Possiamo operare un campionamento con ripetizione estraendo una pallina, osservandola e rimettendola nell'urna dopo aver annotato l'informazione relativa.

In tal caso, se operiamo k estrazioni, avremo

$$\overbrace{nnn\dots n}^{k \text{ volte}} = n^k$$

possibili uscite in quanto per ogni elemento estratto avremo sempre n possibili scelte.

Possiamo anche operare un campionamento senza ripetizione, estraendo, osservando e non rimettendo la pallina nell'urna; in tal caso per la prima estrazione avremo n possibilità, per la seconda $n-1$, per la terza $n-2$ e così via. Pertanto in questo caso avremo

$${}_n D_k = \frac{n!}{(n-k)!}$$

possibili uscite.

4. Le variabili aleatorie

Se è assegnato uno spazio di probabilità $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ diciamo che ξ è una variabile aleatoria definita su $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ se è data una funzione

$$\xi : \mathcal{U} \rightarrow \mathbb{R}$$

Nel caso del lancio di due dadi la terna $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ è definita da

- \mathcal{U} è l'insieme delle 36 coppie di valori (i, j)
- \mathcal{F} è la famiglia di tutti i sottoinsiemi di \mathcal{U}
- \mathcal{P} è definita da $\mathcal{P}(A_{i,j}) = \frac{1}{36}$ ed inoltre se $A \in \mathcal{F}$ è un sottoinsieme costituito da k elementi possiamo definire $\mathcal{P}(A) = k \frac{1}{36}$

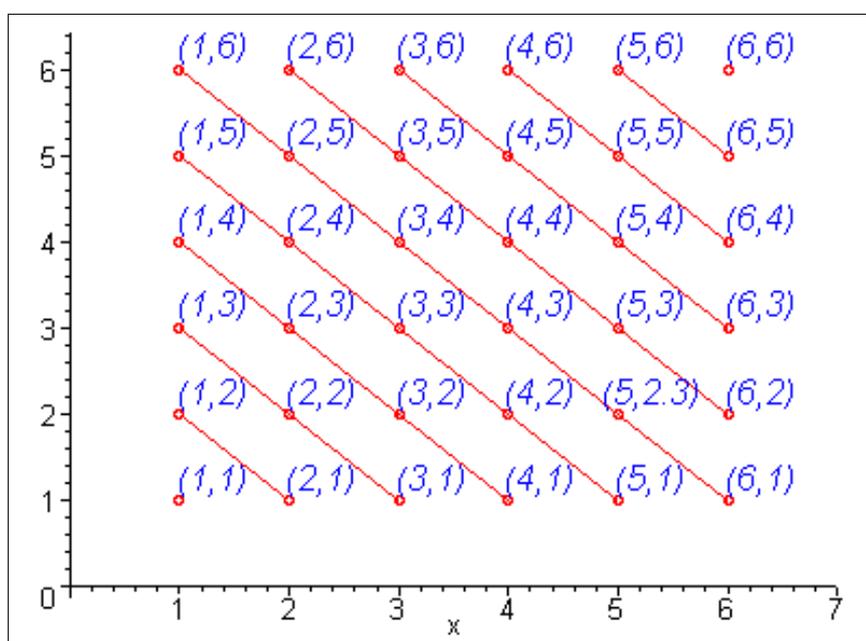


FIGURA 28.3. Eventi che forniscono lo stesso valore per ξ , (congiunti da un segmento)

Possiamo costruire un esempio di variabile aleatoria sullo spazio di probabilità associato al lancio di due dadi assegnando ad ogni uscita $(i, j) \in \mathcal{U}$ il valore ottenuto sommando i punteggi; in altre parole

$$\xi(A_{i,j}) = i + j$$

Possiamo osservare che i punti (eventi) che nella figura 28.3 sono congiunti da un segmento di retta forniscono lo stesso valore per la variabile aleatoria ξ e possiamo riassumere i valori assunti dalla variabile aleatoria ξ nella seguente tabella, dove all'incrocio della riga i -esima e della colonna j -esima possiamo leggere il valore $i + j$ assunto da ξ .

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

TABELLA 28.2. I valori assunti dalla Variabile aleatoria ξ

Possiamo ora definire alcuni elementi caratteristici di una variabile aleatoria:

DEFINIZIONE 28.2. Se ξ è una variabile aleatoria discreta definita su uno spazio di probabilità $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ e se

$$\mathcal{U} = \{A_1, A_2, \dots, A_n\}$$

- la **media** μ di ξ è definita da

$$\mu = \sum_i \xi(A_i) \mathcal{P}(A_i)$$

- la **varianza** σ^2 di ξ è definita da

$$\sigma^2 = \text{Var}(\xi) = \sum_i (\xi(A_i) - \mu)^2 \mathcal{P}(A_i)$$

- lo **scarto quadratico medio** di ξ è definito da

$$\sigma = \sqrt{\sigma^2}$$

- la **moda** M di ξ è definita da

$$M = \xi(A_i)$$

dove A_i è tale che

$$\mathcal{P}(A_i) = \max_j \mathcal{P}(A_j)$$

- la **mediana** m di ξ è definita da

$$\mathcal{P}(\xi \leq m) = \mathcal{P}(\xi \geq m)$$

- il **momento di ordine** k μ_k di ξ è definito da

$$\mu_k = \sum_i (\xi(A_i) - \mu)^k \mathcal{P}(A_i)$$

- *il momento di ordine k , rispetto all'origine μ'_k di ξ è definito da*

$$\mu'_k = \sum_i (\xi(A_i))^k \mathcal{P}(A_i)$$

DEFINIZIONE 28.3. *Chiamiamo funzione di distribuzione di probabilità della variabile aleatoria ξ la funzione*

$$\varphi : \mathcal{U} \rightarrow \mathbb{R}$$

definita da

$$\varphi(i) = \mathcal{P}(\xi = A_i)$$

Se φ è la funzione distribuzione di probabilità di una variabile aleatoria ξ avremo che

$$\mathcal{P}(n \leq \xi \leq m) = \sum_{i=n}^m \varphi(i)$$

$$\mathcal{P}(\xi \leq n) = \sum_{i \leq n} \varphi(i)$$

DEFINIZIONE 28.4. *Se ξ è una variabile aleatoria discreta la cui densità di probabilità è φ , definiamo **speranza matematica di ξ***

$$E(\xi) = \sum_i \xi(A_i) \varphi(i) = \sum_i \xi(A_i) \mathcal{P}(\xi = A_i) = \mu$$

DEFINIZIONE 28.5. *Se ξ è una variabile aleatoria discreta la cui densità di probabilità è φ e se $g : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione, definiamo una nuova variabile aleatoria che indichiamo con $g(\xi)$ mediante la funzione distribuzione di probabilità*

$$\psi(i) = g(\xi(A_i)) \varphi(i)$$

Definiamo inoltre

$$E(g(\xi)) = \sum_i g(\xi(A_i)) \varphi(i) =$$

Osserviamo che si ha

$$\sigma = E((\xi - \mu)^2)$$

$$\mu_k = E((\xi - \mu)^k)$$

$$\mu'_k = E(\xi^k)$$

DEFINIZIONE 28.6. Se ξ è una variabile aleatoria discreta la cui densità di probabilità è φ , definiamo **funzione generatrice dei momenti di ξ** la

$$M_{\xi}(t) = E(e^{t\xi}) = \sum_i e^{t\xi(A_i)} \varphi(i) =$$

4.1. Un esempio. Consideriamo il solito esempio dei dadi e la solita variabile ξ che associa ad ogni uscita in \mathcal{U} la somma dei punteggi, possiamo vedere che

$$\begin{aligned} \varphi(2) &= \mathcal{P}(\xi = 2) = \frac{1}{36} = \mathcal{P}(\xi = 12) = \varphi(12) \\ \varphi(3) &= \mathcal{P}(\xi = 3) = \frac{2}{36} = \mathcal{P}(\xi = 11) = \varphi(11) \\ \varphi(4) &= \mathcal{P}(\xi = 4) = \frac{3}{36} = \mathcal{P}(\xi = 10) = \varphi(10) \\ \varphi(5) &= \mathcal{P}(\xi = 5) = \frac{4}{36} = \mathcal{P}(\xi = 9) = \varphi(9) \\ \varphi(6) &= \mathcal{P}(\xi = 6) = \frac{5}{36} = \mathcal{P}(\xi = 8) = \varphi(8) \\ \varphi(7) &= \mathcal{P}(\xi = 7) = \frac{6}{36} \end{aligned}$$

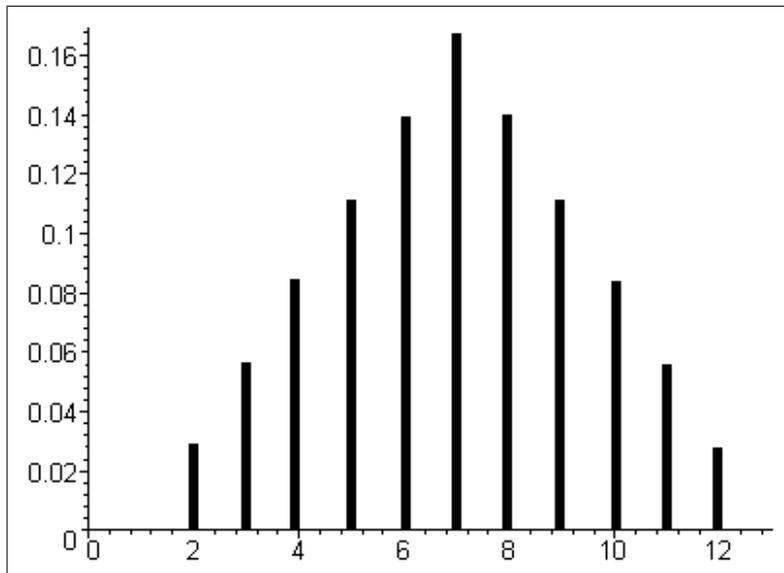


FIGURA 28.4. Distribuzione di probabilità della variabile aleatoria che restituisce il punteggio ottenuto lanciando due dadi.

Possiamo rappresentare questa distribuzione di probabilità mediante il grafico in figura 28.4 (scatter plot) oppure mediante il grafico in figura 28.5.

Quest'ultimo grafico si chiama istogramma ed è quello che più frequentemente viene adottato per rappresentare la funzione di distribuzione di una variabile aleatoria discreta.

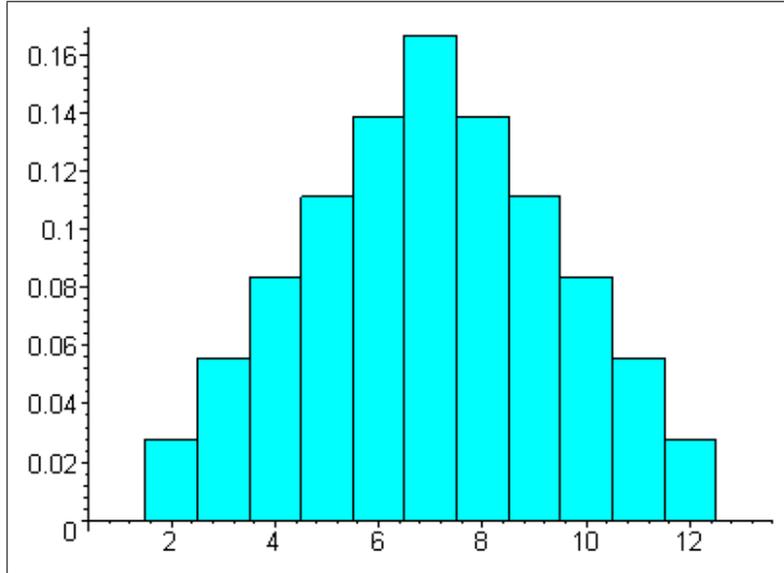


FIGURA 28.5. Istogramma relativo alla variabile aleatoria che restituisce il punteggio ottenuto lanciando due dadi.

Osservando il grafico possiamo osservare che il valore che compare con maggior frequenza è il 7, che pertanto è la moda della variabile ξ .

Possiamo calcolare la probabilità $\mathcal{P}(\xi \leq 7)$ sommando le aree dei rettangoli mostrati in figura(si osservi che la base dei rettangoli è lunga 1) ed otteniamo che $\mathcal{P}(\xi \leq 7) = \frac{21}{36}$.

In modo simile possiamo calcolare $\mathcal{P}(\xi \geq 7) = \frac{21}{36}$ e possiamo concludere che la mediana della variabile aleatoria ξ è $m = 7$

Un semplice calcolo mostra che la varianza di ξ è $\sigma^2 = \frac{35}{6}$

Abbiamo già notato che i rettangoli che costituiscono l'istogramma hanno base lunga 1; ciò permette di definire la funzione di distribuzione di ξ come

$$(28.6) \quad \varphi(x) = \begin{cases} \mathcal{P}(\xi = \eta) & \text{per } \eta - 0.5 \leq x \leq \eta + 0.5 \quad \eta = 2, 3, 4, \dots, 12 \\ 0 & \text{altrove} \end{cases}$$

Con tale definizione si vede che φ è definita su tutto \mathbb{R} ed inoltre si vede che

$$\int_{-\infty}^{+\infty} \varphi(t) dt = \sum_i \varphi(i) = 1$$

essendo la somma estesa a tutti i valori η che variabile discreta ξ può assumere.

5. Variabili aleatorie continue

Talvolta tuttavia non è possibile considerare uno spazio di probabilità discreto, finito o numerabile.

Ciò accade ad esempio quando si considera il problema di scegliere un numero a caso compreso tra 0 ed 1.

Infatti la probabilità di estrarre, ad esempio, il valore 0.3 non si può calcolare considerando il rapporto tra casi favorevoli, uno solo, e casi possibili, infiniti non numerabili.

Anche la definizione di media e varianza presentano qualche problema in quanto occorre definire come si intende procedere per calcolare la somma di un numero infinito, non numerabile, di addendi.

Per chiarire la questione possiamo osservare che, se è difficile definire la probabilità che la variabile aleatoria ξ il cui valore è il numero scelto a caso in $[0, 1]$ assuma il valore x , è invece naturale definire la probabilità che $\xi \in [x, x + h]$.

In tal caso infatti possiamo identificare i casi favorevoli con un segmento di lunghezza h e la totalità dei casi con l'intero intervallo $[0, 1]$ che risulta ovviamente di lunghezza 1.

Pertanto

$$\mathcal{P}(x \leq \xi \leq x + h) = \frac{h}{1}$$

Ricordando il significato di somma dell'integrale, possiamo definire la funzione distribuzione di probabilità della variabile aleatoria ξ come la funzione continua φ tale che

$$(28.7) \quad \mathcal{P}(x \leq \xi \leq x + h) = h = \int_x^{x+h} \varphi(t) dt$$

per ogni $x \in [0, 1]$ e per ogni h abbastanza piccolo.

Ne deduciamo che

$$(28.8) \quad \frac{1}{h} \int_x^{x+h} \varphi(t) dt = 1$$

e, passando al limite per $h \rightarrow 0$, poichè abbiamo supposto φ continua,

$$\varphi(x) = 1$$

Da quanto abbiamo detto appare ragionevole che, nel caso di una variabile aleatoria continua ξ , non è significativo definire

$$(28.9) \quad \mathcal{P}(\xi = x)$$

mentre è naturale definire

$$(28.10) \quad \mathcal{P}(x_0 \leq \xi \leq x_1) = \int_{x_0}^{x_1} \varphi(t) dt$$

dove φ è la funzione di distribuzione di probabilità di ξ .

Pertanto supporremo nota una variabile aleatoria continua ξ se è nota la sua funzione di distribuzione di probabilità φ .

Una funzione $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, continua, è la funzione di distribuzione di probabilità di una variabile aleatoria se

- $\varphi(t) \geq 0$

- **per ogni $t \in \mathbb{R}$**

- $$\int_{-\infty}^{+\infty} \varphi(t) dt = 1$$

In tal caso si ha

$$\mathcal{P}(\xi \leq x) = \int_{-\infty}^x \varphi(t) dt$$

$$\mathcal{P}(x_0 \leq \xi \leq x_1) = \int_{x_0}^{x_1} \varphi(t) dt$$

La funzione

$$F(x) = \mathcal{P}(\xi \leq x) = \int_{-\infty}^x \varphi(t) dt$$

si chiama distribuzione cumulativa di probabilità della variabile aleatoria ξ di densità di probabilità φ .

Osserviamo che ad ogni variabile aleatoria discreta (finita) si può associare una variabile aleatoria continua la cui densità è una funzione costante a tratti, nulla al di fuori di un insieme limitato (nel caso in cui la variabile sia discreta e finita).

Per chiarire il concetto consideriamo la variabile aleatoria ξ che fornisce il punteggio ottenuto nel lancio di due dadi; in tal caso la funzione distribuzione di probabilità φ può essere definita come segue

$$\varphi(x) = \begin{cases} 0 & x < 1.5 \\ k/36 & k + .5 \leq x < k + 1.5 & k = 1..5 \\ 6/36 & 6.5 \leq x < 7.5 \\ (12 - k)/36 & k + .5 \leq x < k + 1.5 & k = 7..11 \\ 0 & x \geq 12.5 \end{cases}$$

Come nel caso delle variabili aleatorie discrete possiamo definire

DEFINIZIONE 28.7. Se ξ è una variabile aleatoria continua che ha densità di probabilità φ ,

- la **media** μ di ξ è definita da

$$\mu = E(\xi) = \int_{-\infty}^{+\infty} x\varphi(x)dx$$

- la **varianza** σ^2 di ξ è definita da

$$\sigma^2 = \text{Var}(\xi) = E((\xi - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 \varphi(x)dx$$

- lo **scarto quadratico medio** di ξ è definito da

$$\sigma = \sqrt{\sigma^2}$$

- la **moda** M di ξ è definita da

$$M = \sup_{x \in \mathbb{R}} \varphi(x)$$

- la **mediana** m di ξ è definita da

$$\mathcal{P}(\xi \leq m) = \int_{-\infty}^m \varphi(x)dx = \int_m^{+\infty} \varphi(x)dx = \mathcal{P}(\xi \geq m)$$

- il **momento di ordine** k μ_k di ξ è definito da

$$\mu_k = E((\xi - \mu)^k) = \int_{-\infty}^{+\infty} (x - \mu)^k \varphi(x)dx$$

- il **momento di ordine** k , **rispetto all'origine** μ'_k di ξ è definito da

$$\mu_k = E(\xi^k) = \int_{-\infty}^{+\infty} x^k \varphi(x)dx$$

DEFINIZIONE 28.8. Se ξ è una variabile aleatoria continua e se $f : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione possiamo definire la variabile aleatoria $f(\xi)$ assegnandone la funzione distribuzione di probabilità definita da

$$\psi(t) = f(t)\varphi(t)$$

In tal modo

$$\mathcal{P}(x_0 \leq f(\xi) \leq x_1) = \int_{x_0}^{x_1} f(t)\varphi(t)dt$$

$$E(f(\xi)) = \int_{-\infty}^{+\infty} f(t)\varphi(t)dt$$

DEFINIZIONE 28.9. Se ξ è una variabile aleatoria continua la cui densità di probabilità è φ , definiamo **funzione generatrice dei momenti** di ξ la

$$M_\xi(t) = E(e^{t\xi}) = \int_{-\infty}^{+\infty} e^{tx}\varphi(x)dx$$

5.1. Le variabili aleatorie discrete infinite. Un caso che si colloca a metà tra quello delle variabili aleatorie discrete finite ed il caso delle variabili aleatorie continue è il caso delle variabili aleatorie discrete che assumono una quantità infinita numerabile di valori.

Il caso delle variabili aleatorie infinite ci suggerisce come comportarci in questo caso.

Se lo spazio \mathcal{U} degli eventi è numerabile, allora potremo scrivere che

$$\mathcal{U} = \{A_i, i \in \mathbb{N}\}$$

e sarà sufficiente definire la probabilità di ciascun evento

$$\mathcal{P}(A_i) = p_i$$

con la condizione

$$\sum_{i=1}^{+\infty} p_i = 1$$

Media, varianza, funzione di distribuzione possono essere definiti mediante le

DEFINIZIONE 28.10. Se ξ è una variabile aleatoria discreta numerabile definiamo

- la **media** μ di ξ come

$$(28.11) \quad \mu = E(\xi) = \sum_{i=1}^{+\infty} i p_i$$

- la **varianza** σ^2 di ξ come

$$(28.12) \quad \sigma^2 = \text{Var}(\xi) = E((\xi - \mu)^2) = \sum_{i=1}^{+\infty} (i - \mu)^2 p_i$$

- lo **scarto quadratico medio** o deviazione standard di ξ

$$\sigma = \sqrt{\sigma^2}$$

- il **momento k -esimo** μ_k di ξ come

$$\mu_k = \sum_{i=1}^{+\infty} (i - \mu)^k p_i$$

- il **momento k -esimo rispetto all'origine** μ'_k di ξ come

$$\mu'_k = \sum_{i=1}^{+\infty} i^k p_i$$

- la **funzione di distribuzione** φ di ξ come

$$(28.13) \quad \varphi(n) = \mathcal{P}(\xi = A_i)$$

e si ha

$$(28.14) \quad \mathcal{P}(\xi \leq n) = \sum_{i=1}^n \mathcal{P}(\xi = A_i) \quad n \in \mathbb{N}$$

Naturalmente possiamo estendere la definizione di φ ad \mathbb{R} definendola costante sugli intervalli del tipo $[k, k+1]$ e ponendola uguale a 0 prima di 1.

Inoltre

DEFINIZIONE 28.11. Se ξ è una variabile aleatoria discreta infinita la cui densità di probabilità è φ , e se $f : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione possiamo definire la variabile aleatoria $f(\xi)$ assegnandone la funzione distribuzione di probabilità definita da

$$\psi(k) = f(k)\varphi(k) = f(k)p_k$$

Quindi

$$E(f(\xi)) = \sum_{i=1}^{+\infty} f(i)\varphi(i)$$

Definiamo anche definiamo **funzione generatrice dei momenti di ξ** la

$$M_{\xi}(t) = E(e^{t\xi}) = \sum_{k=1}^{+\infty} e^{kt}\varphi(k) =$$

6. Distribuzioni di probabilità doppie

Siano $(\mathcal{U}_1, \mathcal{F}_1, \mathcal{P}_1)$ e $(\mathcal{U}_2, \mathcal{F}_2, \mathcal{P}_2)$ due spazi di probabilità consideriamo la variabile aleatoria che indichiamo con (ξ, η) definita sullo spazio $\mathcal{U}_1 \times \mathcal{U}_2$ mediante la

$$F(x, y) = \mathcal{P}(\xi \leq x, \eta \leq y) = \int_{-\infty}^x \left(\int_{-\infty}^y f(t, s) ds \right) dt$$

F è la distribuzione cumulativa di probabilità della variabile (ξ, η) ed f è la sua funzione distribuzione di probabilità

Se f è continua possiamo affermare che

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y)$$

Naturalmente devono essere verificate le seguenti condizioni:

•

$$f(x, y) \geq 0$$

•

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(t, s) ds \right) dt = 1$$

Inoltre se

$$F_1(x) = \mathcal{P}(\xi \leq x) = \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} f(t, s) ds \right) dt$$

$$F_2(y) = \mathcal{P}(\eta \leq y) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^y f(t, s) ds \right) dt = \int_{-\infty}^y \left(\int_{-\infty}^{+\infty} f(t, s) dt \right) ds$$

F_1 ed F_2 sono le distribuzioni cumulative delle variabili aleatorie ξ e η , rispettivamente le cui funzioni di distribuzione sono date da

$$\varphi(t) = \int_{-\infty}^{+\infty} f(t, s) ds$$

$$\psi(s) = \int_{-\infty}^{+\infty} f(t, s) dt$$

Nel caso in cui le variabili aleatorie ξ e η siano indipendenti, allora (ξ, η) ha una distribuzione di probabilità

$$f(t, s) = \varphi(t)\psi(s)$$

dove φ e ψ sono le funzioni di distribuzione di ξ e η , rispettivamente.

È utile ricordare che la probabilità della variabile aleatoria ξ condizionata alla variabile aleatoria η si può definire mediante la

$$\mathcal{P}(\xi \leq x | \eta \leq y) = \int_{-\infty}^x \left(\int_{-\infty}^y \frac{f(t, s)}{\psi(s)} ds \right) dt$$

per cui $\frac{f(t, s)}{\psi(s)}$ è la sua funzione di distribuzione di probabilità.

7. Qualche proprietà di valor medio e varianza

Abbiamo già definito cosa intendiamo per

- media o valor medio
- varianza
- momento
- momento rispetto all'origine
- speranza matematica
- funzione generatrice dei momenti

di una variabile aleatoria ξ sia nel caso discreto finito sia nel caso continuo sia nel caso discreto infinito.

Tra le proprietà del valor medio ricordiamo che

- $E(\alpha\xi + \beta\eta) = \alpha E(\xi) + \beta E(\eta)$
- **se ξ e η sono variabili aleatorie indipendenti**

$$E(\xi\eta) = E(\xi)E(\eta)$$

In altre parole la speranza matematica è una applicazione lineare; La verifica di tale fatto non è banale e si può ad esempio ottenere provando il risultato nel caso discreto finito ed utilizzando un passaggio al limite per gli altri casi.

Si può provare che anche la varianza gode di proprietà di linearità, infatti

- $\text{Var}(\alpha\xi) = \alpha^2 \text{Var}(\xi)$
- **se ξ e η sono variabili aleatorie indipendenti**

$$\text{Var}(\xi \pm \eta) = \text{Var}(\xi) + \text{Var}(\eta)$$

È anche utile ricordare che

$$\begin{aligned} \sigma^2 &= E((\xi - \mu)^2) = E(\xi^2 - 2\mu\xi + \mu^2) = \\ &= E(\xi^2) - 2\mu E(\xi) + \mu^2 = \\ &= E(\xi^2) - 2\mu^2 + \mu^2 = E(\xi^2) - \mu^2 = E(\xi^2) - (E(\xi))^2 \end{aligned}$$

Vediamo ora come la funzione generatrice dei momenti si riveli molto comoda per il calcolo dei momenti di una variabile aleatoria.

Cominciamo con l'osservare che

$$(t - \mu)^k = \sum_{i=0}^k \binom{k}{i} t^i \mu^{k-i}$$

moltiplicando per $\varphi(t)$ ed integrando, otteniamo

$$(28.15) \quad \mu_k = \sum_{i=0}^k \binom{k}{i} \mu^i \mu^{k-i}$$

e se ne ricava che per trovare i momenti rispetto al valor medio μ_k è sufficiente conoscere i momenti rispetto all'origine μ'_k .

Casi particolari della 28.15 sono

$$\begin{aligned} \mu_2 &= \sigma_2 = \mu'_2 - \mu^2 \\ \mu_3 &= \mu'_3 - 3\mu'_2\mu + 2\mu^3 \end{aligned}$$

(ricordiamo che $\mu_0 = \mu'_0 = 1$ e $\mu_1 = \mu'_1 = \mu$).

D'altro canto i momenti di una variabile aleatoria si possono trovare a partire dalla funzione generatrice dei momenti che è definita da:

$$M_{\xi}(t) = E(e^{t\xi}) = \int_{-\infty}^{+\infty} e^{ts} \varphi(s) ds$$

Infatti si può verificare che la funzione M_{ξ} è sviluppabile in serie di McLaurin ed il suo sviluppo è dato da

$$M_{\xi}(t) = \sum_{i=0}^{+\infty} \mu'_i \frac{t^i}{i!}$$

e quindi

$$\mu'_k = \frac{d^k}{dt^k} M_{\xi}(t)$$

8. La disuguaglianza di Tchebichev e la legge dei grandi numeri

In questa sezione ci occupiamo di due risultati fondamentali: la disuguaglianza di Tchebichev e la legge dei grandi numeri, cominciando a parlare della prima.

Sia ξ una variabile aleatoria con media μ e varianza σ^2 , allora si avrà che

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{+\infty} (t - \mu)^2 \varphi(t) dt = \\ &= \int_{\{t : |t - \mu| \leq \varepsilon\}} (t - \mu)^2 \varphi(t) dt + \int_{\{t : |t - \mu| > \varepsilon\}} (t - \mu)^2 \varphi(t) dt \geq \\ &\geq \int_{\{t : |t - \mu| \geq \varepsilon\}} (t - \mu)^2 \varphi(t) dt \geq \int_{\{t : |t - \mu| \geq \varepsilon\}} \varepsilon^2 \varphi(t) dt = \\ &= \varepsilon^2 \mathcal{P}(|\xi - \mu| \geq \varepsilon) \end{aligned}$$

se ne ricava pertanto che

$$(28.16) \quad \mathcal{P}(|\xi - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

La 28.16 è nota come disuguaglianza di Tchebichev e ne possiamo trarre una interessante conseguenza: per $\varepsilon = k\sigma$ otteniamo che

$$(28.17) \quad \mathcal{P}(|\xi - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Pertanto

$$(28.18) \quad \mathcal{P}(|\xi - \mu| < k\sigma) = 1 - \mathcal{P}(|\xi - \mu| \geq k\sigma) \geq 1 - \frac{1}{k^2}$$

Se ora consideriamo la seguente tabella

k	$1 - \frac{1}{k^2}$
1	0
2	.75
3	.88
4	.93
5	.95
6	.97

TABELLA 28.3. Valori approssimati di $1 - \frac{1}{k^2}$

Si vede pertanto che se ξ è una variabile aleatoria di media μ e di varianza σ^2 , allora la probabilità che il valore assunto da ξ sia vicino alla media μ per meno di 2 volte la varianza è del 75% e sale all'88% se ci accontentiamo di un errore inferiore a 3 volte la varianza.

Va osservato che, nonostante fornisca risultati soddisfacenti, la disuguaglianza di Tchebichev non è molto precisa.

8.1. La legge dei grandi numeri. Un'altra delle conseguenze della disuguaglianza di Tchebichev prende il nome di "Legge dei grandi numeri" e si ricava come segue.

Siano $\xi_1, \xi_2, \dots, \xi_n$ variabili aleatorie tutte con media μ e varianza σ^2 , e consideriamo la variabile aleatoria

$$S_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}$$

Avremo che

$$E(S_n) = \frac{1}{n}(E(\xi_1) + E(\xi_2) + \dots + E(\xi_n)) = \mu$$

ed inoltre

$$\text{Var}(S_n) = \frac{1}{n^2}(\text{Var}(\xi_1) + \text{Var}(\xi_2) + \dots + \text{Var}(\xi_n)) = \frac{\sigma^2}{n}$$

Pertanto

$$(28.19) \quad \mathcal{P}(|S_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$$

per $n \rightarrow +\infty$

La 28.19 è nota con il nome di **legge dei grandi numeri** ed esprime un concetto in base al quale la media delle uscite di una variabile aleatoria differisce dalla media della variabile aleatoria di una quantità infinitesima.

Va tuttavia sottolineato che la legge dei grandi numeri fornisce informazioni di carattere qualitativo e quindi non può essere usata per stime di tipo quantitativo.

9. Normalizzazione di una variabile aleatoria.

Sia ξ una variabile aleatoria di media μ e di varianza σ^2 con distribuzione di probabilità φ .

Sia cioè

$$\begin{aligned}\int_{-\infty}^{+\infty} \varphi(t) dt &= 1 \\ \mathcal{P}(a \leq \xi \leq b) &= \int_a^b \varphi(t) dt \\ E(\xi) &= \int_{-\infty}^{+\infty} t\varphi(t) dt = \mu \\ E(\xi^2) &= \int_{-\infty}^{+\infty} (t - \mu)^2 \varphi(t) dt = \sigma^2\end{aligned}$$

e consideriamo la variabile aleatoria

$$\xi^* = \frac{\xi - \mu}{\sigma}$$

Per le proprietà di media e varianza possiamo affermare che ξ^* è una variabile normalizzata (o standardizzata), intendendo con ciò che ξ^* ha media 0 e varianza 1.

Allo scopo di determinare la funzione di distribuzione di ξ^* consideriamo la variabile aleatoria η la cui funzione di distribuzione è definita da

$$\sigma\varphi(\mu + \sigma t)$$

Possiamo allora verificare che

$$\begin{aligned}\int_{-\infty}^{+\infty} \sigma\varphi(\mu + \sigma t) dt &= \int_{-\infty}^{+\infty} \varphi(s) ds = 1 \\ \int_{-\infty}^{+\infty} t\sigma\varphi(\mu + \sigma t) dt &= \int_{-\infty}^{+\infty} \left(\frac{s - \mu}{\sigma}\right) \varphi(s) ds = \\ &= \frac{1}{\sigma} \left(\int_{-\infty}^{+\infty} s\varphi(s) ds - \mu \int_{-\infty}^{+\infty} \varphi(s) ds \right) = 0 \\ \int_{-\infty}^{+\infty} t^2\sigma\varphi(\mu + \sigma t) dt &= \int_{-\infty}^{+\infty} \left(\frac{s - \mu}{\sigma}\right)^2 \varphi(s) ds = \\ &= \frac{1}{\sigma} \int_{-\infty}^{+\infty} (s - \mu)^2 \varphi(s) ds = 1\end{aligned}$$

$$\begin{aligned}\mathcal{P}(a \leq \eta \leq b) &= \int_a^b \sigma \varphi(\mu + \sigma t) dt = \\ &= \int_{\mu + \sigma a}^{\mu + \sigma b} \sigma \varphi(s) \frac{1}{\sigma} ds = \mathcal{P}(\mu + \sigma a \leq \xi \leq \mu + \sigma b) = \\ &= \mathcal{P}\left(a \leq \frac{\xi - \mu}{\sigma} \leq b\right)\end{aligned}$$

Ne deduciamo che η e $\xi^* = \frac{\xi - \mu}{\sigma}$ hanno la stessa distribuzione di probabilità e quindi sono la stessa variabile aleatoria.

In altre parole la funzione

$$\sigma \varphi(\mu + \sigma t)$$

è la distribuzione di probabilità di

$$\xi^* = \frac{\xi - \mu}{\sigma}$$

QUALCHE DISTRIBUZIONE DI PROBABILITÀ

Le funzioni di distribuzione di probabilità sono fondamentali per descrivere il comportamento delle variabili aleatorie che ci interessano.

Ogni variabile aleatoria ha una sua distribuzione e per definirne la proprietà è utile fare riferimento ad alcune distribuzioni note che sono in grado di descrivere la maggior parte delle variabili aleatorie con cui normalmente si lavora.

1. La distribuzione uniforme

La più semplice funzione di distribuzione di probabilità è quella di una variabile aleatoria che restituisce un valore scelto in un intervallo $[a, b]$ con il criterio di equiprobabilità.

Abbiamo già visto che in tal caso

$$\mathcal{P}(x \leq \xi \leq x + h) = \frac{h}{b - a}$$

e che la sua distribuzione di densità è

$$\varphi(t) = \begin{cases} \frac{1}{b-a} & t \in [a, b] \\ 0 & \text{altrove} \end{cases}$$

La funzione generatrice dei momenti si calcola mediante la

$$M_{\xi}(t) = \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Se ne ricava subito che

$$\mu = \frac{b+a}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

Descriviamo ora tre importanti distribuzioni di probabilità che sono, come vedremo, tra loro legate:

- La distribuzione binomiale di Bernoulli
- La distribuzione di Poisson
- La distribuzione normale di Gauss

Si tratta di una distribuzione discreta finita (Bernoulli), una distribuzione discreta numerabile (Poisson) ed una distribuzione continua (Gauss) che possono essere derivate a partire dal concetto elementare di prova bernoulliana.

2. La distribuzione binomiale di Bernoulli

DEFINIZIONE 29.1. Chiamiamo prova bernoulliana un esperimento che ha due soli possibili esiti:

- Successo, cui associamo il valore 1 con probabilità p
- Insuccesso, cui associamo il valore 0 con probabilità q

essendo ovviamente $p + q = 1$.

Chiamiamo variabile aleatoria bernoulliana la variabile aleatoria ξ che restituisce il numero di successi che si sono verificati su n prove ripetute (lanci) dell'esperimento.

Possiamo calcolare la probabilità che la variabile aleatoria ξ assuma il valore k mediante la

$$\varphi(k) = \mathcal{P}(\xi = k) = \binom{n}{k} p^k q^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

Possiamo giustificare la formula precedente se descriviamo la successione di n prove ripetute con una stringa di elementi che assumono il valore 1 oppure 0 a seconda che la corrispondente prova abbia avuto o no successo.

0 1 1 0 0 1 0 0 1 0 0 1 0 1 0 0 0 1 0 1

affinchè ci siano k successi la stringa dovrà contenere esattamente k volte il valore 1 (ed $n - k$ volte il valore 0) e quindi, poichè in ogni elemento 1 si presenta con probabilità p e 0 con probabilità q , una stringa con k successi avrà una probabilità di comparire uguale a

$$p^k q^{n-k}$$

d'altro canto, poichè siamo unicamente interessati a contare il numero di successi, e non l'ordine con cui si verificano, dovremo tener conto che si verificano, ad esempio, 2 successi in tanti modi diversi

0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

il cui numero è dato dalle combinazioni di n oggetti a k a k e cioè

$$\binom{n}{k}$$

(Infatti ciascuna stringa può essere individuata dalla sequenza dei k numeri, compresi tra 1 ed n , che indicano la posizione dei successi nella stringa.)

Possiamo calcolare la media della variabile bernoulliana ξ osservando che la media in ciascuna prova è

$$1 \cdot p + 0 \cdot q = p$$

e su n esperimenti, essendo la media lineare, avremo

$$\mu = E(\xi) = np$$

La varianza della variabile bernoulliana ξ in ciascuna prova è

$$(1-p)^2 \cdot p + 0 \cdot (0-q)^2 q = q^2 p + p^2 q = pq(p+q) = pq$$

e su n esperimenti essendo la varianza lineare avremo

$$\sigma^2 = E((\xi - \mu)^2) = npq$$

e

$$\sigma = \sqrt{npq}$$

Per calcolare la funzione generatrice dei momenti possiamo procedere come segue

$$\begin{aligned} M_\xi(t) &= E(e^{t\xi}) = \sum_{k=0}^n e^{tk} \binom{n}{k} p^k q^{n-k} = \\ &= \sum_{k=0}^n \binom{n}{k} (pe^t)^k q^{n-k} = (pe^t + q)^n \end{aligned}$$

Un esempio notevole di prove bernoulliane si costruisce considerando il lancio di una moneta ripetuto per n volte; in tal caso $p = q = \frac{1}{2}$.

3. La distribuzione di Poisson

Si consideri il caso di m prove ripetute bernoulliane in cui la probabilità di successo della singola prova sia $\frac{1}{n}$

Un esempio di tale situazione di si può realizzare considerando un pagliaio con un numero $n - 1$ di pagliuzze dal quale si voglia estrarre 1 ago; la probabilità di estrarre l'ago sarà ovviamente $\frac{1}{n}$ e la prova sarà ripetuta m volte.

Sia ξ la variabile aleatoria che restituisce il numero di successi ottenuti nelle m prove fatte. Avremo che

$$\begin{aligned} \varphi(k) &= \mathcal{P}(\xi = k) = \binom{m}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k} = \\ &= \frac{m(m-1)(m-2)\cdots(m-(k-1))}{k!} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{-k} = \\ &= \frac{1}{k!} \frac{m}{n} \left(\frac{m-1}{n}\right) \cdots \left(\frac{m-k+1}{n}\right) \left(1 - \frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{-k} \end{aligned}$$

Se ora consideriamo di far tendere n a $+\infty$ mantenendo

$$\frac{m}{n} = \lambda$$

(oppure chiedendo semplicemente che $\frac{m}{n} \rightarrow \lambda$), otteniamo che

$$\varphi(k) = \frac{1}{k!} \lambda \left(\lambda - \frac{1}{n} \right) \left(\lambda - \frac{2}{n} \right) \cdots \left(\lambda - \frac{k-1}{n} \right) \left(1 - \frac{1}{n} \right)^{n\lambda} \left(1 - \frac{1}{n} \right)^{-k}$$

e, per $n \rightarrow +\infty$

$$\varphi(k) = \frac{1}{k!} \lambda^k e^{-\lambda}$$

La funzione distribuzione di probabilità di Poisson è quindi data da

$$\boxed{\varphi(k) = \frac{1}{k!} \lambda^k e^{-\lambda}}$$

Per calcolare media, varianza ed i momenti della distribuzione di Poisson è utile calcolare la funzione generatrice dei momenti.

$$M_{\xi}(t) = \sum_{k=0}^{+\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{+\infty} \frac{(\lambda e^t)^k}{k!} e^{-\lambda} = e^{-\lambda} e^{\lambda(e^t-1)}$$

e le sue derivate

$$\begin{aligned} \frac{d}{dt} M_{\xi}(t) &= e^{\lambda(e^t-1)} \lambda e^t \\ \frac{d^2}{dt^2} M_{\xi}(t) &= e^{\lambda(e^t-1)} \lambda^2 e^t + e^{\lambda(e^t-1)} \lambda e^t \end{aligned}$$

e calcolarle in $t = 0$.

$$\begin{aligned} \frac{d}{dt} M_{\xi}(0) &= \lambda = \mu'_1 = \mu \\ \frac{d^2}{dt^2} M_{\xi}(0) &= \lambda^2 + \lambda = \mu'_2 \end{aligned}$$

da cui

$$\begin{aligned} \mu &= \mu'_1 = \lambda \\ \sigma^2 &= \mu'_2 - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

4. La distribuzione gaussiana

Sia $\xi_n = k$ una variabile aleatoria bernoulliana relativa ad n prove ripetute, allora

$$\mathcal{P}(\xi_n = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Sia ha

$$\begin{aligned} E(\xi_n) &= \mu = np \\ Var(\xi_n) &= \sigma^2 = npq \end{aligned}$$

Consideriamo la variabile normalizzata (di media 0 e varianza 1)

$$\xi_n^* = \frac{\xi_n - \mu}{\sigma}$$

ξ_n^* assume i valori

$$x_k = \frac{k - \mu}{\sigma} \quad k = 1, \dots, n$$

e

$$\mathcal{P}\left(\xi_n^* = \frac{k - \mu}{\sigma}\right) = \mathcal{P}(\xi_n = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Possiamo ricavare inoltre che

$$\begin{aligned} k &= \mu + \sigma x_k = np + \sigma x_k \\ n - k &= n - np - \sigma x_k = nq - \sigma x_k \end{aligned}$$

dove $\sigma = \sqrt{npq}$ e si vede che, per $n \rightarrow +\infty$ tanto k quanto $n - k$ tendono a $+\infty$, inoltre

$$\begin{aligned} \frac{k}{np} &= 1 + \sqrt{\frac{q}{np}} x_k \\ \frac{n-k}{nq} &= 1 + \sqrt{\frac{p}{nq}} x_k \\ x_{k+1} - x_k &= \Delta x_k = \frac{1}{\sigma} = \frac{1}{\sqrt{npq}} \end{aligned}$$

Possiamo allora scrivere che

$$\begin{aligned} \mathcal{P}(x' \leq \xi_n^* \leq x'') &= \sum_{x_k \in [x', x'']} \mathcal{P}(\xi_n^* = x_k) = \\ &= \sum_{\{k: x_k \in [x', x'']\}} \mathcal{P}(\xi_n = k) = \\ &= \sum_{\{k: x_k \in [x', x'']\}} \frac{n!}{k!(n-k)!} p^k q^{n-k} \end{aligned}$$

Poichè, per n grande, tanto k quanto $n - k$ sono grandi, possiamo usare la formula di Stirling per approssimare il fattoriale e quindi

$$\begin{aligned} \frac{n!}{k!(n-k)!} p^k q^{n-k} &\approx \frac{n^n e^{-n} \sqrt{2\pi n}}{k^k e^{-k} \sqrt{2\pi k} (n-k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n-k)}} p^k q^{n-k} = \\ &= \frac{1}{\sqrt{2\pi}} \frac{n^k}{k^k} \frac{n^{n-k}}{(n-k)^{n-k}} \sqrt{\frac{n}{k(n-k)}} p^k q^{n-k} = \\ &= \frac{1}{\sqrt{2\pi}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \sqrt{\frac{n}{k(n-k)}} \end{aligned}$$

Ma, sempre per n grande,

$$\sqrt{\frac{n}{k(n-k)}} = \sqrt{\frac{n}{(np + \sigma x_k)(nq - \sigma x_k)}} \approx \sqrt{\frac{n}{n^2 pq}} = \frac{1}{\sqrt{npq}}$$

D'altro canto

$$\begin{aligned}
\ln \left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} &= -k \ln \left(\frac{np}{k} \right) - (n-k) \ln \left(\frac{nq}{n-k} \right) = \\
&= -(np + \sigma x_k) \ln \left(1 + \sqrt{\frac{q}{np}} \right) - \\
&\quad - (nq - \sigma x_k) \ln \left(1 - \sqrt{\frac{p}{nq}} x_k \right) + \omega \left(\frac{1}{\sqrt{n}} \right) = \\
&= -(np + \sigma x_k) \left(\sqrt{\frac{q}{np}} x_k - \frac{1}{2} \frac{q}{np} x_k^2 \right) - \\
&\quad - (nq - \sigma x_k) \left(-\sqrt{\frac{p}{nq}} x_k - \frac{1}{2} \frac{p}{nq} x_k^2 \right) + \omega \left(\frac{1}{\sqrt{n}} \right) = \\
&= -x_k \left[np \sqrt{\frac{q}{np}} - nq \sqrt{\frac{p}{nq}} \right] - x_k^2 \left[\sqrt{npq} \sqrt{\frac{q}{np}} + \sqrt{npq} \sqrt{\frac{p}{nq}} \right] - \\
&\quad - x_k^2 \left[\frac{1}{2} np \frac{q}{np} + \frac{1}{2} nq \frac{p}{nq} \right] - x_k^3 \left[\frac{q \sqrt{npq}}{2np} + \frac{p \sqrt{npq}}{nq} \right] = \\
&= -x_k^2 + \frac{1}{2} x_k^2 + \omega \left(\frac{1}{\sqrt{n}} \right)
\end{aligned}$$

dove, come al solito, si indica con ω una funzione infinitesima.

Ne concludiamo pertanto che

$$\begin{aligned}
\mathcal{P}(x' \leq \xi_n^* \leq x'') &\approx \\
&\approx \sum_{x_k \in [x', x'']} \frac{1}{\sqrt{2\pi}} \left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} \sqrt{\frac{n}{k(n-k)}} \approx \\
&\approx \sum_{x_k \in [x', x'']} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Delta x_k \rightarrow \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-\frac{1}{2}t^2} dt
\end{aligned}$$

È pertanto naturale considerare la funzione distribuzione di probabilità definita da

$$\boxed{\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}}$$

La φ si chiama funzione di distribuzione di probabilità Gaussiana di media 0 e varianza 1. (Si ricordi che è ottenuta come limite di distribuzioni normalizzate).

Utilizzando il teorema di integrazione per sostituzione (poichè l'integrale è improprio deve essere calcolato come limite di integrali propri ed a questi si applica il teorema di integrazione per sostituzione) possiamo ricavare che la funzione

$$\varphi(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

è una funzione di distribuzione di probabilità, che chiameremo ancora Gaussiana, di media μ e di varianza σ^2 .

Possiamo calcolare la funzione generatrice dei momenti della funzione di distribuzione gaussiana, infatti

$$M_\xi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

e, operando il cambio di variabili $\frac{x-\mu}{\sigma} = u$, da cui $x = \sigma u + \mu$,

$$\begin{aligned} M_\xi(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t\sigma u + t\mu} e^{-\frac{u^2}{2}} \sigma du = \\ &= \sigma \frac{1}{\sigma\sqrt{2\pi}} e^{t\mu} \int_{-\infty}^{+\infty} e^{t\sigma u} e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} e^{t\mu} \int_{-\infty}^{+\infty} e^{\frac{1}{2}(2t\sigma u - u^2)} du = \\ &= \frac{1}{\sqrt{2\pi}} e^{t\mu} \int_{-\infty}^{+\infty} e^{\frac{1}{2}(\sigma^2 t^2 - \sigma^2 t^2 + 2t\sigma u - u^2)} du = \\ &= \frac{1}{\sqrt{2\pi}} e^{t\mu + \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(t\sigma - u)^2} du = \\ &= \frac{1}{\sqrt{2\pi}} e^{t\mu + \frac{1}{2}\sigma^2 t^2} \sqrt{2\pi} = e^{t\mu + \frac{1}{2}\sigma^2 t^2} \end{aligned}$$

5. Due distribuzioni discrete

5.1. La distribuzione multinomiale. Consideriamo un esperimento che possa avere k possibili esiti, che indichiamo con

$$A_1, A_2, \dots, A_k$$

con probabilità

$$p_1, p_2, \dots, p_k \quad , \quad p_1 + p_2 + \dots + p_k = 1$$

e supponiamo di replicarlo per n volte; consideriamo la variabile aleatoria ξ che restituisce la n -pla di valori

$$n_1, n_2, \dots, n_k \quad , \quad n_1 + n_2 + \dots + n_k = n$$

dove n_i è il numero di volte in cui si è verificato l'evento A_i .

La funzione distribuzione di probabilità di ξ è data da

$$\begin{aligned}\varphi(n_1, n_2, \dots, n_k) &= \mathcal{P}(\xi_1 = p_1, \xi_2 = p_2, \dots, \xi_k = p_k) = \\ &= \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}\end{aligned}$$

5.2. La distribuzione ipergeometrica. Consideriamo un'urna contenente b palline nere e w palline bianche e supponiamo di estrarre per n volte una pallina rimettendola. dopo ogni estrazione, nell'urna.

Consideriamo la variabile aleatoria ξ che restituisce il numero di volte in cui si è estratta una pallina nera; allora la densità di probabilità di ξ si può calcolare mediante la

$$\begin{aligned}\varphi(k) = \mathcal{P}(\xi = k) &= \binom{n}{k} \left(\frac{b}{b+w}\right)^k \left(\frac{w}{b+w}\right)^{n-k} = \\ &= \binom{n}{k} \frac{b^k w^{n-k}}{b+w}\end{aligned}$$

non appena si ricordi la distribuzione binomiale e si tenga presente che

$$p = \frac{b}{b+w}, \quad q = \frac{w}{b+w}$$

Qualora l'esperimento si ripeta senza rimettere la pallina estratta nell'urna, (campionamento senza ripetizione), si può vedere che la densità di probabilità della nuova variabile aleatoria ξ che conta il numero delle palline nere estratte è

$$\varphi(k) = \mathcal{P}(\xi = k) = \frac{\binom{b}{k} \binom{w}{n-k}}{\binom{b+w}{n}}$$

Si calcola anche che

$$\begin{aligned}\mu &= \frac{nb}{b+w} \\ \sigma^2 &= \frac{nbw(b+w-n)}{(b+w)^2(b+w-1)}\end{aligned}$$

6. Le distribuzioni legate ai test statistici.

6.1. La distribuzione χ^2 . Si tratta della distribuzione di probabilità di una variabile aleatoria χ^2 che restituisce la somma di ν variabili aleatorie ξ_i indipendenti, aventi distribuzione gaussiana con media 0 e varianza 1 (distribuzioni normali standardizzate).

$$\chi^2 = \xi_1^2 + \xi_2^2 + \dots + \xi_\nu^2$$

La funzione di distribuzione della variabile aleatoria χ^2 è data da

$$\varphi(t) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} t^{\nu/2-1} e^{-t/2} & t \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

ν rappresenta i gradi di libertà, media, varianza e generatrice dei momenti sono date da

$$\begin{aligned} \mu &= \nu \\ \sigma^2 &= 2\nu \\ M_{\chi^2}^2(t) &= (1 - 2t)^{-\nu/2} \end{aligned}$$

6.2. La distribuzione T di Student. È la distribuzione di una variabile aleatoria T che restituisce il rapporto

$$T = \frac{\xi}{\sqrt{\eta/\nu}}$$

dove ξ è una variabile aleatoria con densità di probabilità gaussiana normale (media 0 e varianza 1) ed η è una variabile aleatoria con distribuzione χ^2 a ν gradi di libertà

La funzione di distribuzione della variabile aleatoria T è data da

$$\varphi(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

ν rappresenta i gradi di libertà, media e varianza sono date da

$$\begin{aligned} \mu &= 0 \\ \sigma^2 &= \frac{\nu}{\nu - 2} \end{aligned}$$

6.3. La distribuzione F di Fisher. È la distribuzione di una variabile aleatoria F che restituisce il rapporto

$$F = \frac{\xi_1/\nu_1}{\xi_2/\nu_2}$$

dove ξ_1 ed ξ_2 sono variabili aleatorie con distribuzione χ^2 a ν_1 e ν_2 gradi di libertà, rispettivamente.

La funzione di distribuzione della variabile aleatoria F è data da

$$\varphi(t) = \begin{cases} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \nu_1^{\nu_1/2} \nu_2^{\nu_2/2} t^{\nu_2/2-1} (\nu_2 + \nu_1 t)^{-(\nu_1+\nu_2)/2} & t > 0 \\ 0 & \text{altrimenti} \end{cases}$$

media e varianza sono date da

$$\mu = \frac{\nu_2}{\nu_2 - 2}$$

$$\sigma^2 = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)(\nu_2 - 2)^2}$$

7. La distribuzione γ .

La distribuzione γ è definita, per $\alpha, \beta > 0$ da

$$\varphi(t) = \begin{cases} t^{\alpha-1} e^{-t/\beta} & t > 0 \\ 0 & \text{altrimenti} \end{cases}$$

media, varianza e generatrice dei momenti sono date da

$$\mu = \alpha\beta$$

$$\sigma^2 = \alpha\beta^2$$

$$M_\gamma(t) = (1 - \beta t)^{-\alpha}$$

8. La distribuzione β .

La distribuzione β è definita da

$$\varphi(t) = \begin{cases} \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)} & 0 < t < 1 \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1} & 0 < t < 1 \\ 0 & \text{altrimenti} \end{cases}$$

media e varianza sono date da

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)}$$

Si vede anche la sua moda è $\frac{\alpha-1}{\alpha+\beta-2}$.

9. Variabili casuali con distribuzione assegnata.

Qualora sia necessario utilizzare dati generati casualmente con funzione distribuzione di probabilità fissata, possiamo procedere come segue.

Sia ϕ la distribuzione che si vuole considerare e sia

$$F(x) = \int_{-\infty}^x \phi(t) dt$$

la funzione di distribuzione cumulativa; allora si ha

$$\mathcal{P}(a \leq \xi \leq b) = \int_a^b \phi(t) dt = F(b) - F(a)$$

e quindi

$$\mathcal{P}(F^{-1}(a) \leq \xi \leq F^{-1}(b)) = b - a$$

e

$$\mathcal{P}(a \leq F(\xi) \leq b) = b - a$$

Pertanto $F(\xi)$ ha una distribuzione uniforme e quindi poichè

$$\xi = F^{-1}(F(\xi))$$

possiamo generare valori distribuiti con densità di probabilità ϕ , considerando valori generati con densità uniforme ed applicando a tali valori F^{-1} .

Tale procedimento non è tuttavia applicabile, ad esempio, per determinare valori distribuiti con densità gaussiana in quanto non è possibile determinare esplicitamente F^{-1} nel caso in cui

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

In tal caso, che peraltro è di rilevante importanza possiamo osservare che se ξ e η sono variabili aleatorie indipendenti con distribuzione gaussiana normale, allora (ξ, η) è una variabile aleatoria la cui funzione distribuzione di probabilità è

$$\frac{1}{2\pi} e^{-(t^2+s^2)/2}$$

Pertanto

$$\mathcal{P}((\xi, \eta) \in A) = \frac{1}{2\pi} \iint_A e^{-(t^2+s^2)/2} dt ds = \frac{1}{2\pi} \iint_B \rho e^{-(\rho^2)/2} d\rho d\theta$$

dove B ed A sono l'uno il trasformato dell'altro rispetto al cambio di variabili in coordinate polari.

Ne viene che possiamo identificare due nuove variabili (R, Θ) la cui densità di probabilità è data da

$$\frac{1}{2\pi} \rho e^{-(\rho^2)/2} = \left(\frac{1}{2\pi}\right) (\rho e^{-(\rho^2)/2}) = f(\theta)g(\rho)$$

dove

$$f(\theta) = \frac{1}{2\pi}$$

$$g(\rho) = \rho e^{-(\rho^2)/2}$$

Quindi per quanto visto in precedenza, per generare valori casuali di Θ e di R possiamo utilizzare valori uniformemente distribuiti θ ed r e applicare

a tali valori le funzioni F^{-1} e G^{-1} , rispettivamente, dove

$$F(t) = \int_0^t f(s)ds = 2\pi t \quad , \quad G(t) = \int_0^t g(s)ds = 1 - e^{-t^2/2}$$

Si ha allora

$$F^{-1}(s) = \frac{s}{2\pi} \quad , \quad G^{-1}(s) = \sqrt{-2 \ln(1-s)}$$

e le variabili

$$\xi = \sqrt{-2 \ln(1-t)} \cos\left(\frac{s}{2\pi}\right) \quad , \quad \eta = \sqrt{-2 \ln(1-t)} \sin\left(\frac{s}{2\pi}\right)$$

dove s e t sono distribuite uniformemente, risultano distribuite con densità gaussiana di media 0 e di varianza 1.

10. I test statistici

Consideriamo il seguente esempio.

Supponiamo di voler determinare se una moneta è truccata lanciandola un numero N di volte ed osservando il risultato dei lanci.

Possiamo definire truccata una moneta in cui la probabilità di uscita di T , che chiameremo p è diversa dalla probabilità di uscita di C che è $q = (1-p)$. Sarà quindi truccata una moneta per la quale $p \neq 0.5$.

Osserviamo comunque che allo scopo di ottenere benefici da una scommessa potremmo anche essere interessati a conoscere se $p > 0.5$ o equivalentemente se $p < 0.5$.

I dati di cui vogliamo individuare qualche caratteristica, la popolazione (il nome deriva dalle origini dello studio statistico), sono costituiti da tutti i possibili lanci della moneta in questione.

I dati di cui disponiamo sono costituiti dalle uscite degli N lanci osservati, campione.

Del campione siamo in grado di conoscere la distribuzione di probabilità che è una distribuzione discreta binomiale la cui media è $\mu = Np$ e la cui varianza è $\sigma^2 = \sqrt{Npq}$.

Calcoliamo la probabilità che la variabile aleatoria ξ che restituisce il numero di teste uscite negli N lanci sia compresa tra n ed m .

Avremo che

$$\mathcal{P}(n \leq \xi \leq m) = \sum_{k=n}^m \binom{N}{k} p^k q^{N-k}$$

e quindi, se ad esempio $N = 100$, $n = 45, 40, 30$, $m = 55, 60, 70$,

$$\mathcal{P}(45 \leq \xi \leq 55) \approx 0.73$$

$$\mathcal{P}(40 \leq \xi \leq 60) \approx 0.96$$

$$\mathcal{P}(30 \leq \xi \leq 70) \approx 0.99$$

Possiamo quindi affermare che

- nel caso il numero di teste uscite sia superiore a 55 o inferiore a 45, si è verificato un evento che, supponendo la moneta non truccata, ha la probabilità di verificarsi del $100 - 73 = 27\%$
- nel caso il numero di teste uscite sia superiore a 60 o inferiore a 40, si è verificato un evento che, supponendo la moneta non truccata, ha la probabilità di verificarsi del $100 - 96 = 4\%$
- nel caso il numero di teste uscite sia superiore a 70 o inferiore a 30, si è verificato un evento che, supponendo la moneta non truccata, ha la probabilità di verificarsi del $100 - 99 = 1\%$

e potremmo in base a queste considerazioni concludere che

- nel caso il numero di teste uscite sia superiore a 55 o inferiore a 45, siamo fiduciosi che la moneta sia truccata con un livello di significatività del 73%
- nel caso il numero di teste uscite sia sia superiore a 60 o inferiore a 40, siamo fiduciosi che la moneta sia truccata con un livello di significatività del 96%
- nel caso il numero di teste uscite sia superiore a 70 o inferiore a 30, siamo fiduciosi che la moneta sia truccata con un livello di significatività del 99%

Abbiamo in altre parole usato la variabile ξ che restituisce il numero di teste uscite per stimare la veridicità dell'affermazione:

La moneta non è truccata

Chiamiamo la variabile ξ stimatore e l'affermazione "La moneta non è truccata" ipotesi H_0

Osserviamo che accettare che la moneta sia truccata, cioè rigettare l'ipotesi H_0 che "La moneta non è truccata" con un livello di significatività del 99% non vuol dire essere certi che la moneta è truccata con il 99% di possibilità in quanto non è tra l'altro affatto chiaro rispetto a quale spazio tale probabilità sia calcolata.

Potremmo anche calcolare la probabilità che il numero di teste uscito sia compreso tra n ed m utilizzando il fatto che la distribuzione binomiale si può approssimare con la distribuzione normale.

In tal caso dobbiamo normalizzare i dati osservando che

$$\begin{aligned} \mathcal{P}(n \leq \xi \leq m) &= \mathcal{P}\left(\frac{n - Np}{\sqrt{Npq}} \leq \frac{\xi - Np}{\sqrt{Npq}} \leq \frac{m - Np}{\sqrt{Npq}}\right) = \\ &= G\left(\frac{m - Np}{\sqrt{Npq}}\right) - G\left(\frac{n - Np}{\sqrt{Npq}}\right) \end{aligned}$$

dove

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

è la funzione distribuzione cumulativa normale standardizzata (gaussiana).

Otterremo i seguenti risultati che differiscono di poco dai precedenti

$$\mathcal{P}(45 \leq \xi \leq 55) \approx 0.68$$

$$\mathcal{P}(40 \leq \xi \leq 60) \approx 0.95$$

$$\mathcal{P}(30 \leq \xi \leq 70) \approx 0.99$$

e potremmo trarre conclusioni simili alle precedenti.

Qualora fossimo interessati a verificare l'ipotesi H_0 a fronte dell'ipotesi H_1 che la probabilità di successo p sia inferiore a 0.5, potremmo calcolare

$$\mathcal{P}(\xi \leq 55) \approx 0.86$$

$$\mathcal{P}(\xi \leq 60) \approx 0.98$$

$$\mathcal{P}(\xi \leq 70) \approx 0.99$$

e concludere

- nel caso il numero di teste uscite sia superiore a 55, siamo fiduciosi che la moneta sia truccata con $p > .05$ con un livello di significatività del 86%
- nel caso il numero di teste uscite sia superiore a 60, siamo fiduciosi che la moneta sia truccata con $p > .05$ con un livello di significatività del 98%
- nel caso il numero di teste uscite sia superiore a 70, siamo fiduciosi che la moneta sia truccata con $p > .05$ con un livello di significatività del 99%

Questo esempio mostra come sia possibile trarre conclusioni sulle caratteristiche di una popolazione (l'esito di tutti i possibili lanci di una moneta) esaminandone soltanto un campione (l'esito di un numero finito di lanci della moneta) e stimando le caratteristiche mediante la somma degli esiti (1 nel caso di T , successo, 0 in caso contrario).

Per poter procedere occorre precisare meglio i concetti di popolazione, campione e stimatore.

10.1. Popolazioni. Diciamo che è assegnata una popolazione se è assegnato un insieme \mathcal{U} ed una variabile aleatoria ξ definita su \mathcal{U} .

Chiamiamo popolazione l'insieme $\xi(\mathcal{U})$.

Se ad esempio siamo interessati a stimare il diametro di 9 sferette di acciaio contenute in un cuscinetto, possiamo indicare ciascuna sferetta con un indice e considerare la popolazione i cui elementi sono

$$\{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9\}$$

In questo caso \mathcal{U} è costituito dalle 9 sferette

$$\mathcal{U} = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9\}$$

e possiamo considerare la variabile aleatoria ξ che associa ad ogni sferetta il suo diametro $\xi(k) = d_k$ in modo che

$$\xi(\mathcal{U}) = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9\}$$

La variabile ξ descrive la popolazione che stiamo esaminando.

Se fossimo interessati a studiare il diametro delle sferette che costituiscono la produzione mensile di una macchina, sarebbe molto scomodo elencarle tutte e quindi potremmo descrivere la popolazione assegnando per la variabile ξ una funzione di distribuzione, ad esempio gaussiana di medio μ e varianza σ^2 .

Nel caso del lancio della moneta la popolazione è costituita da tutti i lanci.

10.2. Campioni. Data una popolazione definita dalla variabile aleatoria ξ sullo spazio \mathcal{U} diciamo che è assegnato un campione di taglia n se sono assegnate n variabili aleatorie X_1, X_2, \dots, X_n sullo spazio \mathcal{U} . Indichiamo con X la n -pla delle variabili aleatorie X_1, X_2, \dots, X_n che costituiscono il campione.

Ad esempio possiamo considerare

- X_1 come la variabile aleatoria che restituisce il diametro di una sferetta scelta in \mathcal{U}
- X_2 come la variabile aleatoria che restituisce il diametro di una seconda sferetta scelta in \mathcal{U}
- X_k come la variabile aleatoria che restituisce il diametro di una k -esima sferetta scelta in \mathcal{U}

Nel caso del lancio di una moneta il campione è dato dalle N variabili aleatorie che restituiscono gli esiti degli N lanci effettuati.

10.3. Stimatore. Diciamo che è assegnato uno stimatore, o riassunto campionario, se è assegnata una funzione $\phi = \phi(X_1, X_2, \dots, X_n)$ delle n variabili aleatorie X_1, X_2, \dots, X_n che costituiscono il campione.

Possiamo ad esempio considerare uno stimatore del diametro delle sferette considerando la media dei diametri delle n sferette di cui abbiamo esaminato il diametro per costruire il campione.

In tal caso

$$\bar{X} = \phi(X_1, X_2, \dots, X_n) = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n}(d_1 + d_2 + \dots + d_n)$$

Nel caso del lancio di una moneta lo stimatore è costituito dalla somma dei risultati di ciascun lancio, per cui ϕ è il numero di successi ottenuti su n lanci.

Uno stimatore è a sua volta una variabile aleatoria.

Uno stimatore spesso usato è costituito dalla media dei campioni, come visto nell'esempio delle sferette); indicheremo tale stimatore con

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

11. Risultati sulle distribuzioni campionarie

Si consideri una popolazione individuata da una variabile aleatoria ξ su un insieme \mathcal{U}

un campione $X = (X_1, X_2, \dots, X_n)$ di taglia n estratto dalla popolazione individuata da ξ ed \mathcal{U}

Si possono provare alcuni fatti:

11.1. Se $\mu = E(\xi)$ è la media della popolazione ed \bar{X} è la media campionaria

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

allora

- $E(\bar{X}) = \mu_{\bar{X}} = \mu$
- $E((X - \bar{X})^2) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$

11.2. Se ξ è distribuita normalmente con media μ e varianza σ^2 allora \bar{X} è distribuito normalmente con media μ e varianza σ^2/n .

11.3. Se ξ è distribuita con media μ e varianza σ^2 (anche non normalmente) allora

$$\frac{\bar{X} - \mu}{\sigma\sqrt{n}}$$

è asintoticamente ($n \geq 30$) distribuito normalmente con media μ e varianza σ^2/n .

11.4. Se ξ è distribuita normalmente con media μ e se

$$S = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$$

è la varianza campionaria, allora la variabile aleatoria

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

ha una distribuzione di probabilità di Student con $n - 1$ gradi di libertà.

11.5. Se ξ_1, ξ_2 sono due popolazioni normalmente distribuite con varianza σ_1^2, σ_2^2 e se X_1, X_2 sono due campioni di taglia m ed n rispettivamente estratti dalle due popolazioni, allora la variabile aleatoria

$$\frac{(mS_1^2)/((m-1)\sigma_1^2)}{nS_2^2/((n-1)\sigma_2^2)}$$

ha una distribuzione di probabilità di Fisher con $n-1, m-1$ gradi di libertà.

11.6. Supponiamo di disporre di un certo numero di dati in base ai quali la probabilità di accadimento di un evento A_k è x_k e supponiamo di voler verificare se la distribuzione di probabilità di tali eventi ha una funzione densità di probabilità ξ in base alla quale la probabilità attesa degli eventi A_k è ξ_k .

Possiamo considerare che la variabile aleatoria ξ descrive la popolazione degli eventi possibili di cui i valori x_k costituiscono un campione e possiamo considerare come stimatore la variabile aleatoria definita da

$$\chi^2 = \frac{(\xi_1 - x_1)^2}{x_1} + \frac{(\xi_2 - x_2)^2}{x_2} + \dots + \frac{(\xi_n - x_n)^2}{x_n}$$

a proposito della quale possiamo affermare che ha una distribuzione di probabilità χ^2

- $n-1$ gradi di libertà, nel caso in cui della distribuzione ξ siano noti i parametri
- $n-1-m$ gradi di libertà, nel caso in cui della distribuzione ξ si debbano ricavare i parametri per mezzo di m dati dei riassunti campionari.

Per avere informazioni su una variabile aleatoria è necessario conoscere la distribuzione di probabilità.

Possiamo, per analogia con quanto visto in precedenti casi, congetturare che la variabile aleatoria in esame abbia una particolare distribuzione di probabilità, tuttavia avremo successivamente la necessità di verificare se tale congettura è ragionevole.

In altri casi può accadere di aver bisogno di verificare la ragionevolezza di una ipotesi H_0 mediante affermazioni del tipo "se H_0 è vera allora la variabile aleatoria ξ ha una distribuzione di probabilità di un certo tipo (Gaussiana, χ^2, T, F)"

In ogni caso, stabilita la funzione di distribuzione di probabilità che si vuole sottoporre a test, si procede come segue:

- si individua la precisione ϵ con cui si vuole procedere: se ad esempio si fissa $\epsilon = .05$ significa che si vuole essere certi al 95%
- si individuano α, β in modo che

$$\mathcal{P}(\alpha \leq \xi \leq \beta) = .95$$

- si estrae un valore per ξ e se tale valore cade in $[\alpha, \beta]$ si accetta l'ipotesi, in caso contrario si rifiuta.

11.7. Il Test χ^2 . supponiamo di disporre di una serie di dati e di voler verificare se essi sono distribuiti in accordo con una certa ipotesi che chiamiamo H_1 .

Come esempio possiamo adottare il famoso esperimento di Mendel:

Mendel incrociò tra di loro due tipi di piselli ciascuno dei quali aveva due caratteri

- **la forma (Rotondo o angoloso, R oppure a)**
- **il colore (Giallo o verde, G oppure v)**

Egli supponeva che

- **Rotondo e Giallo fossero caratteri dominanti (distinti dall'iniziale Maiuscola)**
- **angoloso e verde fossero caratteri recessivi (iniziale minuscola)**

egli supponeva cioè che dall'incrocio di due piante si ottenesse una terza pianta con i caratteri scelti tra i dominanti delle due che l'avevano generata.

Mendel ottenne 556 piselli che classificò secondo i caratteri appena descritti nella seguente tabella

Caratteri	Frequenza osservata
Rotondo Giallo	315
Rotondo verde	108
angoloso Giallo	101
angoloso verde	32

TABELLA 29.1. Risultati dell'esperimento di Mendel.

I tipi di piselli possibili sono pertanto 4 ed i possibili incroci sono 16 e possono essere elencati nella tabella 29.1 dove accanto alle caratteristiche delle piante A e B incrociate sono riportate le caratteristiche dell'incrocio.

A	+	B	→	C
RG	+	RG	→	RG
RG	+	Rv	→	RG
RG	+	aG	→	RG
RG	+	av	→	RG
Rv	+	RG	→	RG
Rv	+	Rv	→	Rv
Rv	+	aG	→	RG
Rv	+	av	→	Rv
aG	+	RG	→	RG
aG	+	Rv	→	RG

aG	+	aG	→	aG
aG	+	av	→	aG
av	+	RG	→	RG
av	+	Rv	→	Rv
av	+	aG	→	aG
av	+	av	→	av

TABELLA 29.2. Possibili incroci.

Dalla tabella si evince che

- un pisello Rotondo e Giallo si presenta in 9 casi su 16
- un pisello Rotondo e verde si presenta in 3 casi su 16
- un pisello angoloso e Giallo si presenta in 3 casi su 16
- un pisello angoloso e verde si presenta in 1 caso su 16

per cui possiamo dire che

$$\mathcal{P}(RG) = \frac{9}{16}556 = 321.75$$

$$\mathcal{P}(Rv) = \frac{3}{16}556 = 104.75$$

$$\mathcal{P}(RG) = \frac{3}{16}556 = 104.75$$

$$\mathcal{P}(RG) = \frac{1}{16}556 = 34.75$$

e possiamo affiancare alle frequenze osservate nella tabella 29.1 le frequenze previste per ciascun caso:

Caratteri	Frequenza osservata	frequenza prevista	Differenza
	f_o	f_p	$f_o - f_p$
Rotondo Giallo	315	312.75	2.25
Rotondo verde	108	104.25	3.75
angoloso Giallo	101	104.25	-3.25
angoloso verde	32	34.75	-2.75

TABELLA 29.3. Risultati dell'esperimento di Mendel e frequenze previste.

Possiamo ora chiederci se i dati sperimentali confermano l'ipotesi che abbiamo fatto sui caratteri delle due piante:

Rotondo dominante e angoloso recessivo,

Giallo dominante e verde recessivo.

Si può vedere che è lecito supporre che la variabile aleatoria

$$\xi_k = \frac{f_o - f_p}{\sqrt{f_p}}, \quad k = 1, 2, 3, 4$$

segue una distribuzione gaussiana con media 0 e varianza 1.

Pertanto la somma dei quadrati delle 4 variabili aleatorie indicate in tabella 29.3 segue una distribuzione χ^2 con $3 = 4 - 1$ gradi di libertà.

Calcoliamo quindi

$$\chi^2 = \frac{(315 - 312.75)^2}{312.75} + \frac{(108 - 104.25)^2}{104.25} + \frac{(101 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} \approx 0.47$$

e consideriamo la tavola che fornisce i valori di χ^2 con 3 gradi di libertà.

Possiamo osservare che la tavola fornisce l'indicazione che

$$\mathcal{P}(\chi^2 < .352) = .05$$

$$\mathcal{P}(\chi^2 < .584) = .10$$

per cui

$$\mathcal{P}(\chi^2 < .352) < \mathcal{P}(\chi^2 < .47) < \mathcal{P}(\chi^2 < .584)$$

e pertanto possiamo concludere che se gli eventi fossero casuali avremmo trovato un evento che accade con probabilità maggiore del 5% ma minore del 10%, per cui possiamo affermare che l'ipotesi fatta può essere accettata al livello del 90% ma deve essere rifiutata al livello del 95%

Un secondo esempio di applicazione del test χ^2 si trova considerando il seguente problema:

Si supponga di lanciare un dado a forma di tetraedro per 100 volte e di ottenere le seguenti uscite:

1	2	3	4
23	26	24	27

TABELLA 29.4

Ci si pone il problema di stabilire se il dado è non truccato.

La frequenza attesa per ciascuno dei punteggi è $\frac{100}{4} = 25$ e possiamo quindi calcolare

$$\chi^2 = \frac{(23 - 25)^2}{25} + \frac{(26 - 25)^2}{25} + \frac{(24 - 25)^2}{25} + \frac{(27 - 25)^2}{25} \approx .4$$

dalla tabella dei valori di χ^2 con $4 - 1 = 3$ gradi di libertà si ottiene che

$$\mathcal{P}(\chi^2 < .352) = .05$$

$$\mathcal{P}(\chi^2 < .584) = .10$$

per cui

$$\mathcal{P}(\chi^2 < .352) < \mathcal{P}(\chi^2 < .4) < \mathcal{P}(\chi^2 < .584)$$

per cui, come prima, se il dado fosse truccato, avremmo sperimentato un evento la cui probabilità di accadimento è compresa tra il 5% ed il 10%, per cui possiamo affermare che l'ipotesi fatta può essere accettata al livello del 90% ma deve essere rifiutata al livello del 95%

CAPITOLO 30

IL TEOREMA DEL LIMITE CENTRALE

Il teorema del limite centrale è un risultato di grande importanza in quanto sancisce il fatto che la sovrapposizione di un gran numero di variabili aleatorie aventi media e varianza comune conduce ad una variabile con distribuzione normale (gaussiana).

Più precisamente possiamo dire che

Siano

$$\xi_1, \xi_2, \dots, \xi_n$$

variabili aleatorie indipendenti aventi la stessa distribuzione di probabilità con media μ e varianza σ^2 .

Siano

$$\eta_1, \eta_2, \dots, \eta_n$$

le corrispondenti variabili normalizzate

$$\eta_k = \frac{\xi_k - \mu}{\sigma}$$

e consideriamo la variabile aleatoria ζ definita da

$$\zeta_n = \frac{\eta_1 + \eta_2 + \dots + \eta_n}{\sqrt{n}}$$

(Poichè $E(\eta_i) = 0$ avremo che $E(\zeta_n) = nE(\eta_1 + \eta_2 + \dots + \eta_n) = 0$ e inoltre, poichè $\text{Var}(\eta_i) = 1$ avremo che $\text{Var}(\eta_1 + \eta_2 + \dots + \eta_n) = n$ e quindi $\text{Var}(\zeta_n) = 1$)

Si ha

$$\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n - n\mu}{\sqrt{n}\sigma}$$

e si può dimostrare che

$$\lim_{n \rightarrow +\infty} \mathcal{P}(\alpha \leq \zeta_n \leq \beta) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-x^2/2} dx$$

o equivalentemente che

$$\lim_{n \rightarrow +\infty} \mathcal{P}(a \leq \xi_1 + \xi_2 + \dots + \xi_n \leq b) = \frac{1}{\sqrt{2\pi}} \int_{\frac{a-n\mu}{\sqrt{n}\sigma}}^{\frac{b-n\mu}{\sqrt{n}\sigma}} e^{-x^2/2} dx$$

1. Un esempio di applicazione

Dal teorema del limite centrale risulta che moltissime variabili aleatorie possono essere approssimate mediante una variabile che abbia distribuzione, gaussiana.

Se ad esempio consideriamo la variabile aleatoria ξ che restituisce il numero di successi in una serie di N prove bernoulliane con probabilità di successo p e probabilità di insuccesso $q = 1 - p$, possiamo calcolare la probabilità di ottenere un numero di successi compreso tra n ed m utilizzando la distribuzione di Bernoulli mediante la seguente formula

$$\mathcal{P}(n \leq \xi \leq m) = \sum_{k=n}^m \binom{N}{k} p^k q^{N-k}$$

È subito evidente che, se $m - n$ è grande non è facile portare a termine il calcolo; per ovviare a questo inconveniente possiamo allora utilizzare il teorema del limite centrale e sostituire alla distribuzione di Bernoulli una distribuzione Gaussiana.

Procederemo considerando ξ come la somma di N variabili Bernoulliane ξ_i

$$\xi = \xi_1 + \xi_2 + \xi_3 + \dots + \xi_N$$

che assumono ciascuna valori 0 con probabilità q ed 1 con probabilità p .

La media di ciascuna delle ξ_i sarà $\mu = p$ e la loro varianza sarà $\sigma = \sqrt{pq}$, pertanto

$$\begin{aligned} \mathcal{P}(n \leq \xi \leq m) &= \mathcal{P}(n \leq \xi_1 + \xi_2 + \xi_3 + \dots + \xi_N \leq m) \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{\frac{n-Np}{\sqrt{N}\sqrt{pq}}}^{\frac{m-Np}{\sqrt{N}\sqrt{pq}}} e^{-x^2/2} dx = G\left(\frac{m-Np}{\sqrt{N}\sqrt{pq}}\right) - G\left(\frac{n-Np}{\sqrt{N}\sqrt{pq}}\right) \end{aligned}$$

dove

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx$$

è la funzione di distribuzione cumulativa Gaussiana standardizzata i cui valori si trovano tabulati.

Se ad esempio supponiamo di effettuare $N = 100$ lanci di una moneta non truccata per cui $p = q = 1/2$, possiamo valutare la probabilità che si abbiano un numero di Teste compreso tra $n = 45$ e $m = 55$ mediante la

$$\begin{aligned} \mathcal{P}(45 \leq \xi \leq 55) &\approx G\left(\frac{55 - 50}{\sqrt{100}\sqrt{1/4}}\right) - G\left(\frac{45 - 50}{\sqrt{100}\sqrt{1/4}}\right) = \\ &G(1) - G(-1) \approx 0.68 \end{aligned}$$

Abbiamo visto che, se la moneta non è truccata, su 100 lanci possiamo aspettarci un numero di successi compreso tra 45 e 55 con probabilità

del 68% e quindi potremo dire che la probabilità che i successi non siano compresi tra 45 e 55 è del 32%.

Per verificare se la moneta è truccata con un livello di significatività del 32% potremmo allora effettuare una serie di lanci e rifiutare l'ipotesi della moneta non truccata se otteniamo un numero di successi maggiore di 55 o minore di 45.

Qualora volessimo una maggiore precisione, ad esempio del 60% dovremmo avere

$$G\left(\frac{m - Np}{\sqrt{N}\sqrt{pq}}\right) - G\left(\frac{n - Np}{\sqrt{N}\sqrt{pq}}\right) = .40$$

e quindi dovrebbe risultare

$$G\left(\frac{m - Np}{\sqrt{N}\sqrt{pq}}\right) = .5 + .2 = .7 \quad \text{e} \quad G\left(\frac{n - Np}{\sqrt{N}\sqrt{pq}}\right) = .5 - .2 = .3$$

da cui

$$\frac{m - Np}{\sqrt{N}\sqrt{pq}} = .53 \quad \text{e} \quad \frac{n - Np}{\sqrt{N}\sqrt{pq}} = -.53$$

Se ne ricava che

$$m = 50 + 5 * (.53) \approx 52 \quad \quad n = 50 - 5 * (.53) \approx 48$$

REGRESSIONE LINEARE: LA RETTA DEI MINIMI QUADRATI

Siano assegnate n coppie di dati (punti di \mathbb{R}^2)

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

e si consideri il problema di determinare l'equazione di una retta

$$y = ax + b$$

in corrispondenza della quale risulti minima la quantità

$$\epsilon(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

$\epsilon(a, b)$ è una funzione convessa della variabili (a, b) che tende a $+\infty$ per $(a, b) \rightarrow \infty$ e pertanto ammette uno ed un solo punto di minimo assoluto che si può trovare annullando $\nabla\epsilon$.

Per risolvere il problema dovremo pertanto risolvere il sistema definito dalle equazioni

$$\begin{cases} \frac{\partial \epsilon}{\partial a} = \sum_{i=1}^n -2(y_i - ax_i - b)x_i = 0 \\ \frac{\partial \epsilon}{\partial b} = \sum_{i=1}^n -2(y_i - ax_i - b) = 0 \end{cases}$$

Ne viene che

$$\begin{cases} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - b \sum_{i=1}^n 1 = 0 \end{cases}$$

ovvero

$$(31.1) \quad \begin{cases} \sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i = a \sum_{i=1}^n x_i + nb \end{cases}$$

Dalla seconda delle 31.1 si può vedere che

$$nb = \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i$$

ed anche

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

dove

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

indicano la media dei valori x_i ed y_i , rispettivamente.

Dalla prima delle 31.1 si può invece ottenere che

$$\begin{aligned} \sum_{i=1}^n x_i y_i &= a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \\ &= a \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i \left(\frac{\sum_{i=1}^n y_i}{n} - a \frac{\sum_{i=1}^n x_i}{n} \right) \end{aligned}$$

e

$$\begin{aligned} \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} &= a \left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right) \\ n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i &= a \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \end{aligned}$$

ed infine

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

Inoltre

$$\begin{aligned} nb &= \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \left(\frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \right) = \\ &= \frac{1}{n (\sum_{i=1}^n x_i)^2 - (\sum_{i=1}^n x_i)^2} \left[n \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \left(\sum_{i=1}^n x_i \right)^2 \sum_{i=1}^n y_i \right. \\ &\quad \left. - n \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i + \left(\sum_{i=1}^n x_i \right)^2 \sum_{i=1}^n y_i \right] \end{aligned}$$

e se ne conclude che

$$b = \frac{n \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - n \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

Ora, tenendo conto che

$$\begin{aligned}
 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \\
 &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \bar{x} y_i - \sum_{i=1}^n x_i \bar{y} + \sum_{i=1}^n \bar{x} \bar{y} = \\
 &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} - n \bar{x} \bar{y} + n \bar{x} \bar{y} = \\
 &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} = \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}
 \end{aligned}$$

e che

$$\begin{aligned}
 \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n \bar{x} x_i + \sum_{i=1}^n \bar{x}^2 = \\
 &= \sum_{i=1}^n x_i^2 - 2n \bar{x}^2 + n \bar{x}^2 = \sum_{i=1}^n x_i^2 - n \bar{x}^2
 \end{aligned}$$

si ricava che

$$\begin{aligned}
 \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} &= \\
 &= \frac{1}{n} \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \\
 &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - n^2 \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2} = a
 \end{aligned}$$

Pertanto possiamo esprimere a e b mediante le seguenti formule

$$(31.2) \quad \left\{ \begin{array}{l} a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \bar{y} = a \bar{x} + b \end{array} \right.$$

La prima delle due uguaglianze premette di concludere che a è invariante rispetto alla traslazione degli assi: cioè usando $x - x_0$ ed $y - y_0$ in luogo di x ed y il valore di a non cambia.

La stessa trasformazione cambia invece il valore di b , come si vede dalla seconda uguaglianza da cui si vede anche che la retta di regressione passa per il punto di coordinate (\bar{x}, \bar{y}) che è il baricentro dei dati.

Possiamo anche osservare che a meno di operare una traslazione dei dati riportando l'origine degli assi nel baricentro (\bar{x}, \bar{y}) , possiamo supporre che

$$(31.3) \quad \begin{cases} a = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \\ b = 0 \end{cases}$$

Ora, siano

- s_x^2 la varianza dei dati x_i

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

- s_y^2 la varianza dei dati y_i

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

- s_{xy} la covarianza dei dati (x_i, y_i)

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

Possiamo scrivere la retta di regressione nella forma

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x})$$

e se invertiamo il ruolo di x e di y l'equazione diventa

$$x - \bar{x} = \frac{s_{xy}}{s_y^2} (y - \bar{y})$$

Possiamo misurare la correlazione tra i dati utilizzando il coefficiente definito da

$$(31.4) \quad r = \frac{s_{xy}}{s_x s_y}$$

mediante il quale le equazioni delle due rette prima introdotte diventano

$$\frac{y - \bar{y}}{s_y} = r \frac{x - \bar{x}}{s_x}$$

e

$$r \frac{y - \bar{y}}{s_y} = \frac{x - \bar{x}}{s_x}$$

Chiaramente le due rette coincidono soltanto nel caso in cui

$$r^2 = 1 \quad \text{cioè} \quad r = \pm 1$$

e il fatto che questo accada è indice della correlazione dei dati cioè del fatto che i dati si trovano su una retta.

È ragionevole quindi stimare la maggiore o minore correlazione tra i dati confrontando r^2 con 1: più r^2 è vicino ad 1 e più i dati sono da considerarsi linearmente correlati.

Possiamo inoltre misurare la dispersione dei dati attorno alla retta di regressione mediante la

$$S_{y,x} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i^s)^2}{n}}$$

dove

$$y_i^s = ax_i + b$$

Pertanto

$$\begin{aligned} S_{y,x}^2 &= \frac{\sum_{i=1}^n (y_i - ax_i - b)^2}{n} = \sum_{i=1}^n (y_i^2 + a^2x_i^2 + b^2 - 2ax_iy_i - 2by_i + 2abx_i) = \\ &= \sum_{i=1}^n y_i^2 + a^2 \sum_{i=1}^n x_i^2 + nb^2 - 2a \sum_{i=1}^n x_iy_i - 2b \sum_{i=1}^n y_i + 2ab \sum_{i=1}^n x_i = \end{aligned}$$

e per la definizione di a e b

$$\begin{aligned} &= \sum_{i=1}^n y_i^2 + a^2 \left[\frac{1}{a} \left(\sum_{i=1}^n x_iy_i - b \sum_{i=1}^n x_i \right) \right] + nb^2 - \\ &\quad - 2a \sum_{i=1}^n x_iy_i - 2b \sum_{i=1}^n y_i + 2ab \left[\frac{1}{a} \left(\sum_{i=1}^n y_i - nb \right) \right] = \\ &= \sum_{i=1}^n y_i^2 + a \sum_{i=1}^n x_iy_i - ab \sum_{i=1}^n x_i + nb^2 - \\ &\quad - 2a \sum_{i=1}^n x_iy_i - 2b \sum_{i=1}^n y_i + 2b \sum_{i=1}^n y_i - 2nb^2 = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_iy_i - ab \sum_{i=1}^n x_i - nb^2 = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_iy_i - b \left(a \sum_{i=1}^n x_i + nb \right) = \end{aligned}$$

e ancora per la definizione di a e b

$$= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_iy_i - b \sum_{i=1}^n y_i$$

Quindi

$$S_{y,x}^2 = \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i - b \sum_{i=1}^n y_i$$

Si può d'altro canto verificare che

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

infatti

$$\begin{aligned} & \sum_{i=1}^n (y_i - \bar{y})^2 - a \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \\ &= \sum_{i=1}^n y_i^2 - 2\bar{y} \sum_{i=1}^n y_i + n\bar{y}^2 - a \sum_{i=1}^n x_i y_i + a \sum_{i=1}^n x_i y_i + \\ & \quad + a \sum_{i=1}^n \bar{x} y_i - a \sum_{i=1}^n \bar{x} \bar{y} = \\ &= \sum_{i=1}^n y_i^2 - 2n\bar{y}^2 + n\bar{y}^2 - a \sum_{i=1}^n x_i y_i + an\bar{x}\bar{y} + an\bar{x}\bar{y} - an\bar{x}\bar{y} = \\ &= \sum_{i=1}^n y_i^2 - n\bar{y}^2 - a \sum_{i=1}^n x_i y_i + an\bar{x}\bar{y} = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i + n\bar{y}(\bar{y} - a\bar{x}) = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i + n\bar{y}b = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n y_i \end{aligned}$$

Le precedenti considerazioni permettono quindi di affermare che

$$\begin{aligned} S_{y,x}^2 &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i - b \sum_{i=1}^n y_i = \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - a \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \\ &= \frac{ns_y^2 - ans_{xy}}{n} = s_y^2 - as_{xy} \end{aligned}$$

e dal momento che $a = \frac{s_{xy}}{s_x}$

$$= s_y^2 \left(1 - \frac{s_{xy}}{s_y^2} \right) = s_y^2 \left(1 - \frac{s_{xy}^2}{s_x^2 s_y^2} \right) = s_y^2 (1 - r^2)$$

Ne viene quindi che

$$\frac{S_{y,x}}{s_y^2} = (1 - r^2)$$

e

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i^s)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

D'altro canto

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (y_i - y_i^s + y_i^s - \bar{y})^2 = \\ &= \sum_{i=1}^n (y_i - y_i^s)^2 + \sum_{i=1}^n (y_i^s - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - y_i^s)(y_i^s - \bar{y}) \end{aligned}$$

Poichè valgono le equazioni normali 31.1 che definiscono a e b

$$\begin{aligned} \sum_{i=1}^n (y_i - y_i^s)(y_i^s - \bar{y}) &= \sum_{i=1}^n (y_i - ax_i - b)(ax_i + b - \bar{y}) = \\ &= (b - \bar{y}) \sum_{i=1}^n (y_i - ax_i - b) + a \sum_{i=1}^n x_i (y_i - ax_i - b) \\ &= (b - \bar{y}) \left(\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - nb \right) + a \left[\sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i \right] = 0 \end{aligned}$$

avremo

$$\begin{aligned} r^2 &= 1 - \frac{\sum_{i=1}^n (y_i - y_i^s)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - y_i^s)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \\ &= \frac{\sum_{i=1}^n (y_i^s - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{Variazione spiegata}}{\text{Variazione totale}} \end{aligned}$$

Possiamo anche calcolare, dalla 31.4, che

$$\begin{aligned}
 r &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2\right)}} = \\
 &= \frac{\bar{x}\bar{y} - \bar{x}\bar{y}}{\sqrt{(\bar{x}^2 - \bar{x}^2)(\bar{y}^2 - \bar{y}^2)}} = \\
 &= \frac{n \sum_{i=1}^n x'_i y'_i - \sum_{i=1}^n x'_i \sum_{i=1}^n y'_i}{\sqrt{\left(n \sum_{i=1}^n x_i'^2 - \left(\sum_{i=1}^n x'_i\right)^2\right) \left(n \sum_{i=1}^n y_i'^2 - \left(\sum_{i=1}^n y'_i\right)^2\right)}} = \\
 &= \frac{\sum_{i=1}^n x'_i y'_i}{\sqrt{\left(\sum_{i=1}^n x_i'^2\right) \left(\sum_{i=1}^n y_i'^2\right)}}
 \end{aligned}$$

ANALISI DEI COMPONENTI PRINCIPALI.

Quando si esaminano i dati statistici relativi ad un elevato numero di caratteristiche di una popolazione è opportuno considerare i dati in un sistema di riferimento che ne metta in evidenza le caratteristiche salienti.

Supponiamo ad esempio di campionare una funzione f su un intervallo $[a, b]$; dividiamo cioè l'intervallo in $2n - 1$ parti uguali ed indichiamo con

$$a = t_1, t_2, \dots, t_{2n} = b$$

i punti mediante i quali tale suddivisione è operata.

Consideriamo poi n coppie di dati ottenute come segue

$$P_1 = (f(t_1), f(t_2)), P_2 = (f(t_3), f(t_4)), \dots, P_n = (f(t_{2n-1}), f(t_{2n}))$$

È ragionevole che, se la funzione è continua, ci siano piccole differenze tra ascissa e ordinata di ognuno dei punti P_k e pertanto essi tenderanno a disporsi lungo la bisettrice primo-terzo quadrante $y = x$.

Quindi potremo ottenere una rappresentazione più significativa dei dati se useremo come assi coordinati le bisettrici $y = x$ ed $y = -x$, cioè se opereremo una rotazione di 45° degli assi.

Ricordiamo che se $P = (x, y)$ sono le coordinate originali, si opera una rotazione degli assi applicando una trasformazione lineare la cui matrice di rappresentazione è

$$A = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

Pertanto

$$AP_1 = \left(\frac{\sqrt{2}}{2} f(t_1) + \frac{\sqrt{2}}{2} f(t_2), -\frac{\sqrt{2}}{2} f(t_2) + \frac{\sqrt{2}}{2} f(t_1) \right)$$

ed è evidente che delle due coordinate è rilevante solo la prima, mentre la seconda è piccola.

Ne viene che si possono ricostruire i valori originali con sufficiente precisione anche se le seconde coordinate dei punti trasformati sono trascurate.

Per illustrare il metodo occorre premettere qualche considerazione sulle forme quadratiche,

1. Forme quadratiche ed autovalori.

Sia A una matrice $n \times n$ e consideriamo la funzione

$$f(u) = \langle Au, u \rangle \quad , u \in \mathbb{R}^n$$

f si chiama forma quadratica su \mathbb{R}^n e si può vedere che è sempre possibile supporre che la matrice A che la individua sia simmetrica.

Infatti, ad esempio per $n = 2$, se

$$A_1 = \begin{pmatrix} a & d_1 \\ d_2 & c \end{pmatrix}$$

ed $u = \begin{pmatrix} x \\ y \end{pmatrix}$ avremo che

$$\begin{aligned} f(u) = \langle A_1 u, u \rangle &= ax^2 + d_1 xy + d_2 yx + cy^2 = ax^2 + (d_1 + d_2)xy + cy^2 = \\ &= ax^2 + 2bxy + cy^2 = \langle Au, u \rangle \end{aligned}$$

per $b = (d_1 + d_2)/2$, da cui

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

D'altro canto, se A è una matrice simmetrica possiamo verificare che

$$\langle Au, v \rangle = \langle u, Av \rangle$$

ed inoltre

$$f(u+h) = \langle A(u+h), (u+h) \rangle = \langle Au, u \rangle + 2\langle Au, h \rangle - \langle Ah, h \rangle$$

per cui

$$f(u+h) - f(u) = 2\langle Au, h \rangle + \langle Ah, h \rangle = 2\langle Au, u \rangle + \|h\|\omega(h)$$

con ω funzione infinitesima per $h \rightarrow 0$ (si ricordi che $|\langle Ah, h \rangle| \leq \|Ah\|\|h\|$ ed $Ah \rightarrow 0$ se $h \rightarrow 0$).

Dalla definizione di differenziale si ottiene allora che f è differenziabile e che

$$\nabla f(u) = 2Au$$

Come caso particolare, se $A = I$ (la matrice identica) si ha

$$g(u) = \langle u, u \rangle = \|u\|^2 \quad , \quad \nabla g(u) = 2u$$

Consideriamo ora il problema di trovare

$$\max_{g(u)-1=0} f(u) = \max_{\|u\|^2-1=0} \langle Au, u \rangle$$

Per il teorema di Weierstraß, dal momento che f è continua e che $\|u\|^2 - 1 = 0$ definisce la superficie della sfera di centro l'origine e raggio 1, che è

chiusa e limitata, possiamo affermare che il massimo esiste. Sia u_1 il punto in cui tale massimo è assunto

$$\langle Au_1, u_1 \rangle = \max_{\|u\|^2=1} \langle Au, u \rangle$$

$$\|u_1\|^2 = 1$$

D'altro canto, il teorema dei moltiplicatori di Lagrange consente di affermare che esiste λ_1 tale che

$$\nabla f(u_1) = \lambda_1 \nabla g(u_1)$$

per cui deve essere

$$Au_1 = \lambda_1 u_1$$

Dal momento che la precedente equazione è soddisfatta

- λ_1 è un autovalore di A
- u_1 è un autovettore di A corrispondente all'autovalore λ_1

Possiamo inoltre osservare che

$$\max_{\|u\|^2=1} \langle Au, u \rangle = \langle Au_1, u_1 \rangle = \langle \lambda_1 u_1, u_1 \rangle = \lambda_1 \|u_1\|^2 = \lambda_1$$

per cui λ_1 è il valore massimo assunto da $\langle Au, u \rangle$ sulla sfera $\|u\|^2 = 1$

Consideriamo ora lo spazio vettoriale V_1 generato da u_1

$$V_1 = \{ \lambda u_1 : \lambda \in \mathbb{R} \}$$

e lo spazio V_1^\perp ortogonale a V_1

$$V_1^\perp = \{ v \in \mathbb{R}^n : \langle v, u_1 \rangle = 0 \}$$

V_1^\perp è uno spazio vettoriale $(n-1)$ -dimensionale e la matrice A definisce una forma quadratica f_1 su V_1^\perp in quanto

per ogni $v \in V_1^\perp$ si ha

$$\langle Av, u_1 \rangle = \langle v, Au_1 \rangle = \langle v, \lambda_1 u_1 \rangle = 0$$

e $Av \in V_1^\perp$

Se ora ripetiamo il procedimento per f_1 su V_1^\perp possiamo provare un nuovo autovalore λ_2 ed un nuovo autovettore u_2 con le stesse caratteristiche di λ_1 e u_1 .

Chiaramente si può iterare fino a trovare:

- n autovalori $\lambda_1, \lambda_2, \dots, \lambda_n$ decrescenti in valore
- n autovettori u_1, u_2, \dots, u_n , uno per ogni autovettore, che risultano ortogonali tra loro.

Gli autovettori u_1, u_2, \dots, u_n formano quindi una base ortonormale con la caratteristica che lungo il primo asse si trova il punto di massimo della forma quadratica $\langle Au, u \rangle$ sulla sfera unitaria in \mathbb{R}^2 , lungo il secondo asse si trova il massimo della forma quadratica $\langle Au, u \rangle$ sulla sfera unitaria in V_1^\perp e così via fino all' n -esimo asse.

2. L'applicazione all'analisi delle componenti principali.

Torniamo ora ai nostri dati $P_k = (u_k^1, u_k^2, \dots, u_k^n) \in \mathbb{R}^n$ e cerchiamo di individuare una combinazione lineare delle componenti

$$Q_k = \alpha_1 u_k^1 + \alpha_2 u_k^2 + \dots + \alpha_n u_k^n$$

in modo che la varianza di Q_k sia massima (massima significatività della variabile).

Definiamo

$$v_i = \text{Var}(u_k^i)$$

$$c_{ij} = \text{Cov}(u_k^i, u_k^j)$$

la varianza delle singole componenti e la covarianza delle componenti a due a due e sia R la **matrice di covarianza** dei dati definita mediante la

$$R = \begin{pmatrix} v_1 & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & v_2 & c_{23} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ c_{n1} & c_{n2} & c_{n3} & \cdots & v_n \end{pmatrix}$$

Possiamo allora verificare che

$$\text{Var}(Q_k) = \langle Ra, a \rangle$$

dove

$$a = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

$\text{Var}(Q_k)$ è quindi una forma quadratica cui possiamo applicare il metodo visto nella precedente sezione e mediante tale metodo possiamo individuare in ordine decrescente di significatività le componenti dei dati.

2.1. Un esempio. Per illustrare gli effetti del metodo consideriamo i punti del grafico di $\sin(t)$ nell'intervallo $[0, 2\pi]$; suddividiamo l'intervallo in 199 parti uguali, in modo da individuare 200 punti in $[0, 2\pi]$ e calcoliamo i valori assunti da $\sin(t)$ in tali punti.

Le seguenti istruzioni di Matlab producono come risultato un vettore t che contiene i 200 punti in $[0, 2\pi]$, ed i vettori x ed y che contengono i valori assunti da $\sin(6t)$ nei punti pari e dispari rispettivamente

```
clear all;
step=2*pi/199;
t=0:step:2*pi;
xx=0:2*step: 2*pi;
yy=step:2*step: 2*pi;
x=sin(6*xx);
y=sin(6*yy);
```

Possiamo esaminare nel piano la distribuzione dei punti di coordinate $(x(k), y(k))$ con $k = 1, \dots, 100$ mediante le seguenti istruzioni

```
figure(1)
plot(x,y, '+' );
```

`axis([-2 2 -2 2]);`
 che producono la seguente figura 32.1

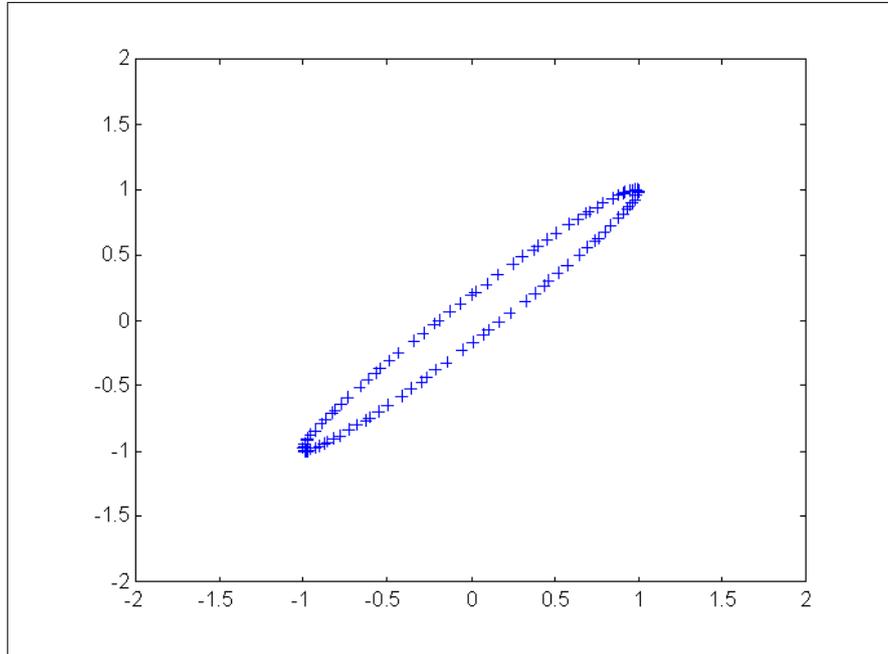


FIGURA 32.1.

Le istruzioni

```
R=cov(x,y);
[V,D]=eig(R)
```

calcolano la matrice di covarianza R e le matrici V e D , dove V è la matrice le cui colonne sono gli autovettori di R e D è una matrice diagonale con gli autovalori sulla diagonale principale; in altre parole V è la matrice tale che

$$VR = RD \quad \text{cioè tale che} \quad V^{-1}RV = D$$

La matrice V pertanto è la matrice di passaggio dal sistema di coordinate originale a quello individuato dagli autovalori di R .

Poichè ci interessa tener conto della componente relativa al massimo autovalore, di assicuriamo anche che l'autovalore più grande sia in posizione (1,1) mediante le istruzioni

```
if D(1,1)>D(2,2)
vv=V(:,1);
V(:,1)=V(:,2);
V(:,2)=vv;
end
```

La matrice V , quindi, può essere usata per effettuare un cambio di base che metta in evidenza le componenti principali.

Le seguenti istruzioni

```
tr=[x(:),y(:)]*V;
```

```
tr1=tr(:,1);
tr2=tr(:,2);
figure(2)
plot(tr1,tr2,'r+')
axis([-2 2 -2 2]);
```

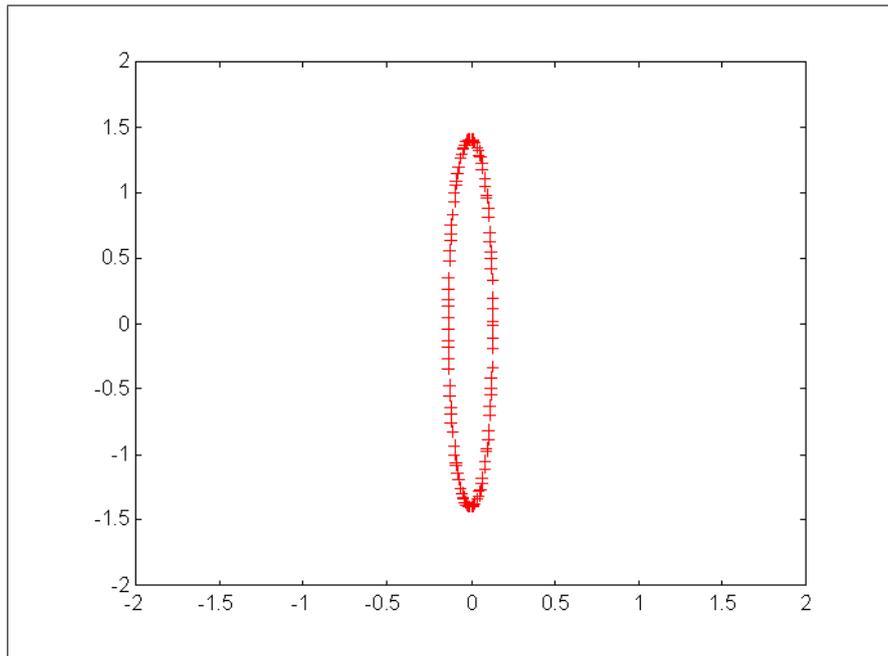


FIGURA 32.2.

Calcolano i trasformati $tr1$, $tr2$ dei punti (x, y) e li mostrano rispetto ad una coppia di assi ortogonali coincidenti con gli autovettori di R (si veda la figura 32.2).

Le successive istruzioni:

```
rtr=[tr1(:),tr2(:)]*inv(V);
rtr1=rtr(:,1);
rtr2=rtr(:,2);
figure(3)
plot(rtr1,rtr2,'g+')
axis([-2 2 -2 2]);
```

mostrano come applicando la trasformazione inversa i dati possano essere recuperati (si veda la figura 32.4.1).

Ora, possiamo osservare che la variazione della seconda componente dei dati trasformati è trascurabile rispetto alla prima, per cui, se la trascuriamo e applichiamo la trasformazione inversa ai dati privati di tale componente, otteniamo nuovi punti che differiscono di poco da quelli originali; possiamo operare usando le seguenti istruzioni:

```
nu=zeros(size(tr2));
ntr=[nu(:),tr2(:)]*inv(V);
```

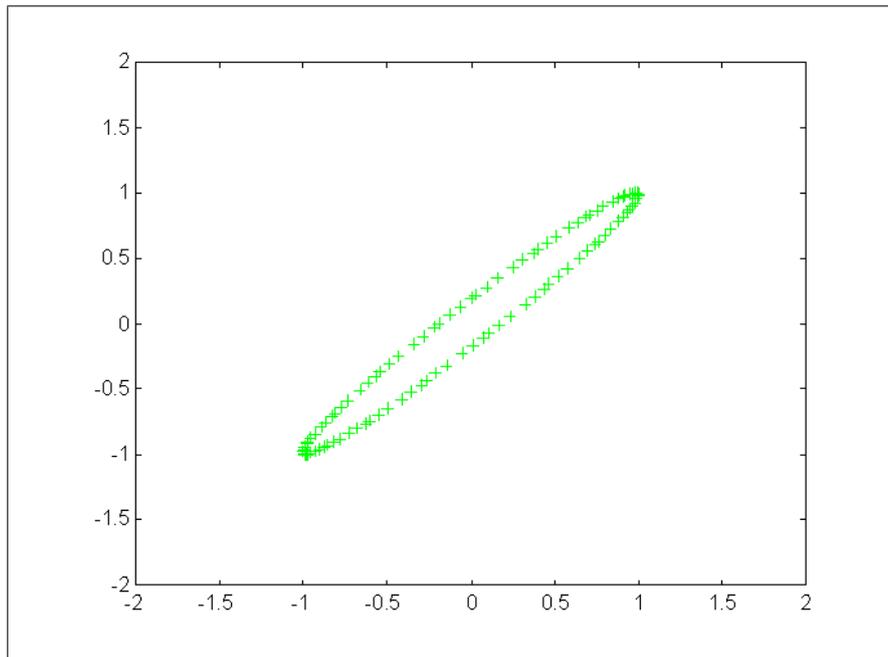


FIGURA 32.3.

```
ntr1=ntr(:,1);
ntr2=ntr(:,2);
figure(4)
plot(ntr1,ntr2,'rx',rtr1,rtr2,'g+')
axis([-2 2 -2 2]);
```

che forniscono anche una immagine dei nuovi punti come indicato in figura 32.4.1 ed un confronto con i punti originali 32.4.2.

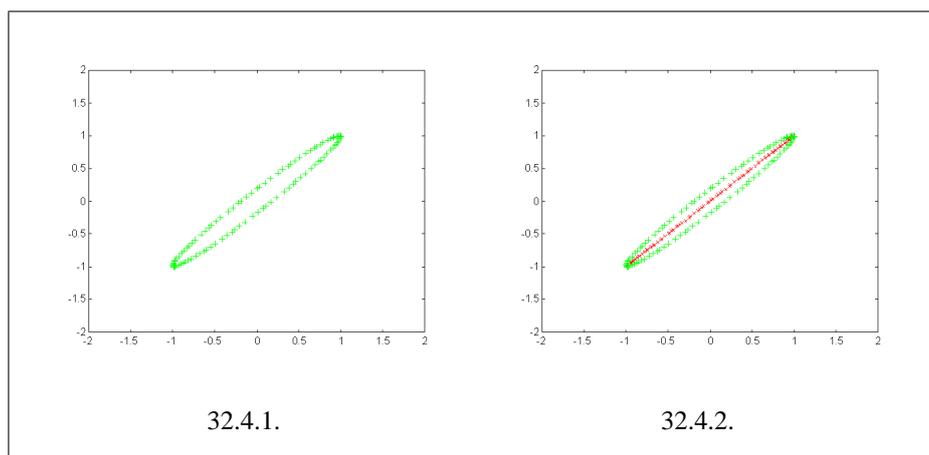


FIGURA 32.4.

È interessante ora osservare come dai punti originali e da quelli privati della componente meno significativa si possa ricostruire la funzione $\sin(6t)$

```

Le seguenti istruzioni
z=zeros(1,200);
zt=zeros(1,200);
for k=0:99
zt(2*k+1)=x(k+1);
end
for k=1:100
zt(2*k)=y(k);
end
for k=0:99
z(2*k+1)=ntr2(k+1);
end
for k=1:100
z(2*k)=ntr1(k);
end
figure(5)
plot(t,z)
figure(6)
plot(t,zt)

```

producono i due grafici riportati in figura 32.5, il primo dei quali riporta la funzione $\sin(t)$ ricostruita congiungendo con segmenti di retta i 200 punti originali, mentre la seconda riporta il grafico di $\sin(6t)$ ricostruito a partire dai punti ottenuti applicando la trasformazione inversa ai punti prima trasformati e poi privati della seconda componente.

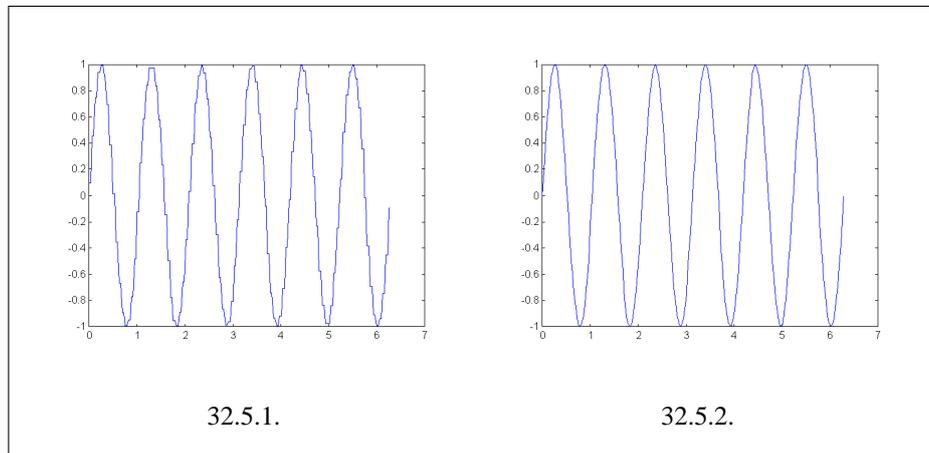


FIGURA 32.5.

Come si vede è evidente che la componente trascurata non ha peggiorato di molto il grafico, mentre la quantità di dati necessari a ricostruire l'immagine si è dimezzata.

Questo indica come può essere sviluppato un procedimento che consenta di immagazzinare dati (i punti del grafico della funzione) utilizzando al meglio le informazioni che contengono.

CAPITOLO 33

Appendice

1. LE FORMULE DI WALLIS E DI STIRLING-DE MOIVRE.

Intendiamo qui provare due formule che sono di grande utilità ed interesse.

La prima, nota come formula del prodotto di Wallis, consente di individuare una successione di razionali che ha per limite π , la seconda, dovuta a Stirling e De Moivre, consente di valutare, con notevole precisione, $n!$.

cominciamo con l'osservare che la definizione di fattoriale consente si affermare che

Si ha

$$(2n)!! = 2^n n! \quad , \quad (2n-1)!! = \frac{(2n)!}{2^n n!}$$

Possiamo provare il seguente teorema

TEOREMA 33.1. -Wallis - Posto

$$w_n = \frac{((2n)!!)^2}{((2n-1)!!)^2 (2n+1)}$$

si ha

$$\lim w_n = \frac{\pi}{2}.$$

Infatti sia

$$s_n = \int_0^{\pi/2} (\sin x)^n dx$$

si ha

$$s_0 = \pi/2 \quad , \quad s_1 = 1$$

ed anche

$$\begin{aligned} s_n &= \int_0^{\pi/2} (\sin x)^{n-1} (\sin x) dx = -(\sin x)^{n-1} \cos x \Big|_0^{\pi/2} + \\ &\quad + \int_0^{\pi/2} (n-1)(\sin x)^{n-2} (\cos x)^2 dx = \\ &= (n-1) \int_0^{\pi/2} ((\sin x)^{n-2} - (\sin x)^n) dx = (n-1)(s_{n-2} - s_n) \end{aligned}$$

onde

$$s_n = \frac{n-1}{n} s_{n-2}, \quad n \geq 2$$

Pertanto

$$\begin{aligned} s_{2n} &= \frac{2n-1}{2n} s_{2n-2} = \frac{2n-1}{2n} \frac{2n-3}{2n-2} s_{2n-4} = \\ &= \frac{2n-1}{2n} \frac{2n-3}{2n-2} \cdots \frac{1}{2} s_0 = \frac{(2n-1)!!}{(2n)!!} \frac{\pi}{2} \end{aligned}$$

Analogamente

$$s_{2n+1} = \frac{2n}{2n+1} \frac{2n-2}{2n-1} \cdots \frac{2}{3} s_1 = \frac{(2n)!!}{(2n+1)!!}.$$

Ora, poiché

$$0 \leq s_n = \int_0^{\pi/2} (\sin x)^n dx \leq \frac{\pi}{2},$$

mentre

$$s_{n+1} = \int_0^{\pi/2} (\sin x)^{n+1} dx \leq \int_0^{\pi/2} (\sin x)^n dx = s_n,$$

si ha che

$$\lim s_n = s \in \mathbb{R}.$$

Ma, come si verifica facilmente

$$w_n = \frac{s_{2n+1}}{s_{2n}} \frac{\pi}{2}$$

e pertanto

$$\lim w_n = \frac{s}{s} \frac{\pi}{2} = \frac{\pi}{2}$$

Se ora osserviamo che si ha

$$w_n = \frac{(2^n n!)^2 (2^n n!)^2}{((2n)!)^2 (2n+1)}$$

da cui

$$v_n = \frac{(n!)^2 2^{2n}}{(2n)! \sqrt{n}} = \sqrt{w_n \frac{(2n+1)}{n}} \longrightarrow \sqrt{\pi}$$

possiamo concludere che

Si ha

$$(33.1) \quad \lim \frac{(n!)^2 2^{2n}}{(2n)! \sqrt{n}} = \sqrt{\pi}$$

Come conseguenza della precedente formula possiamo provare la Formula di Stirling che è di grande utilità per approssimare i valori del fattoriale.

Formula di Stirling-De Moivre**Si ha**

$$\lim \frac{n^n e^{-n} \sqrt{2\pi n}}{n!} = 1$$

da cui

$$n! = n^n e^{-n} \sqrt{2\pi n} \theta_n, \quad \text{con } \lim \theta_n = 1.$$

La verifica della formula di Stirling procede come di seguito illustrato. consideriamo la funzione $1/x$ sull'intervallo $[n, n+1]$; la figura 33.1 mostra come sia vera la disuguaglianza

$$\frac{2}{2n+1} = \frac{1}{n+1/2} \leq \int_n^{n+1} \frac{dx}{x} = \ln \left(1 + \frac{1}{n} \right).$$

Ne segue che

$$1 \leq (n+1/2) \ln(1+1/n)$$

ed

$$e \leq \left(1 + \frac{1}{n} \right)^{n+1/2}.$$

Tenendo conto che \ln è una funzione concava, si ottiene,

$$\begin{aligned} 1 + (\ln 2 + \ln 3 + \dots + \ln(n-1)) + \frac{1}{2} \ln n &\geq \\ &\geq \int_1^{3/2} \ln x dx + \int_{3/2}^{n-1/2} \ln x dx + \int_{n-1/2}^n \ln x dx = \\ &= \int_1^n \ln x dx = n \ln n - n + 1 = \\ &= \ln(n^n e^{-n}) + 1 \end{aligned}$$

da cui

$$\ln \left(\frac{n!}{\sqrt{n}} \right) \geq \ln(n^n e^{-n})$$

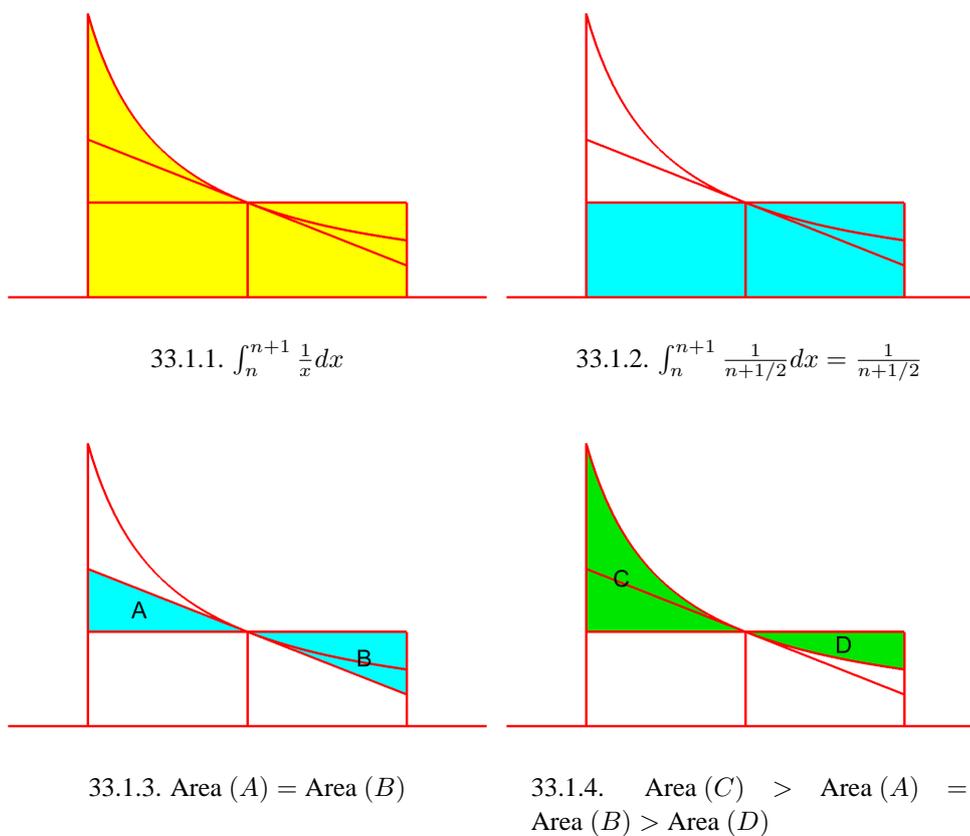


FIGURA 33.1. Grafici di funzioni di due variabili

$$0 \leq u_n = \frac{n^n e^{-n} \sqrt{n}}{n!} \leq 1.$$

Pertanto u_n è limitata e possiamo di più provare che è crescente, verificando che $u_{n+1}/u_n \geq 1$. Si ha infatti

$$\frac{u_{n+1}}{u_n} = \frac{(n+1)^{n+1} e^{-n-1} \sqrt{n+1}}{(n+1)!} \frac{n!}{n^n e^{-n} \sqrt{n}} = \frac{1}{e} \left(1 + \frac{1}{n}\right)^{n+1/2} \geq 1.$$

Possiamo a questo punto affermare che

$$u_n \rightarrow u \in \mathbb{R}$$

e ne consegue che

$$\frac{u_n^2}{u_{2n}} \rightarrow u.$$

Ma

$$\frac{u_n^2}{u_{2n}} = \frac{n^{2n} e^{-2n} n}{(n!)^2} \frac{(2n)!}{(2n)^{2n} e^{-2n} \sqrt{2n}} = \frac{(2n)! \sqrt{n}}{(n!)^2 2^{2n} \sqrt{2}}.$$

Da cui, per la **33.1**

$$u = \lim \frac{u_n^2}{u_{2n}} = \lim \frac{1}{v_n \sqrt{2}} = \frac{1}{\sqrt{2\pi}}$$

e si conclude che

$$\frac{n^n e^{-n} \sqrt{n}}{n!} \longrightarrow \frac{1}{\sqrt{2\pi}}$$

Possiamo anche osservare che,

$\forall n \in \mathbb{N}$

$$.367879 \leq \frac{1}{e} = u_1 \leq u_n \leq \lim u_n = \frac{1}{\sqrt{2\pi}} \leq .398943$$

Elenco delle figure

28.1	Lo spazio \mathcal{U} degli eventi nel caso del lancio di due dadi	4
28.2	Istogramma relativo al lancio di due dadi, (supposti distinti)	5
28.3	Eventi che forniscono lo stesso valore per ξ , (congiunti da un segmento)	14
28.4	Distribuzione di probabilità della variabile aleatoria che restituisce il punteggio ottenuto lanciando due dadi.	17
28.5	Istogramma relativo alla variabile aleatoria che restituisce il punteggio ottenuto lanciando due dadi.	18
32.1		67
32.2		68
32.3		69
32.4		69
32.5		70
33.1	Grafici di funzioni di due variabili	74

Indice

Capitolo 28. ELEMENTI DI PROBABILITÀ E STATISTICA.	3
1. Spazi di probabilità	3
2. Qualche applicazione della regola di Bayes.	8
3. Qualche richiamo di calcolo combinatorio.	11
4. Le variabili aleatorie	14
5. Variabili aleatorie continue	19
6. Distribuzioni di probabilità doppie	23
7. Qualche proprietà di valor medio e varianza	24
8. La disuguaglianza di Tchebichev e la legge dei grandi numeri	26
9. Normalizzazione di una variabile aleatoria.	28
Capitolo 29. QUALCHE DISTRIBUZIONE DI PROBABILITÀ	31
1. La distribuzione uniforme	31
2. La distribuzione binomiale di Bernoulli	32
3. La distribuzione di Poisson	33
4. La distribuzione gaussiana	34
5. Due distribuzioni discrete	37
6. Le distribuzioni legate ai test statistici.	38
7. La distribuzione γ .	40
8. La distribuzione β .	40
9. Variabili casuali con distribuzione assegnata.	40
10. I test statistici	42
11. Risultati sulle distribuzioni campionarie	46
Capitolo 30. IL TEOREMA DEL LIMITE CENTRALE	51
1. Un esempio di applicazione	52
Capitolo 31. REGRESSIONE LINEARE: LA RETTA DEI MINIMI QUADRATI	55
Capitolo 32. ANALISI DEI COMPONENTI PRINCIPALI.	63
1. Forme quadratiche ed autovalori.	64
2. L'applicazione all'analisi delle componenti principali.	66
Capitolo 33. Appendice	71
1. LE FORMULE DI WALLIS E DI STIRLING-DE MOIVRE.	71
Elenco delle figure	77